IRE Transactions on INFORMATION THEORY

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization.

Volume IT-4 DECEMBER, 1958 Number 4

Published Quarterly

In This Issue

PUBLISHED BY THE
Professional Group on Information Theory

IRE Professional Group on Information Theory

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual fee of $3.00.

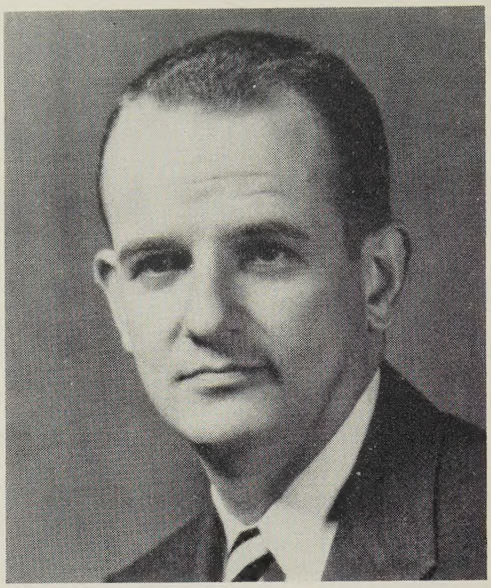
IRE TRANSACTIONS® on INFORMATION THEORY is published by the IRE for the Professional Group on Information Theory, at 1 East 79th Street, New York 21, N. Y. Responsibility for contents rests upon the authors and not upon the IRE, the Group, or its members. Price per copy: IRE-PGIT members, $1.55; IRE members, $2.30; nonmembers, $4.65.

IRE Transactions on Information Theory

Published Quarterly by the Professional Group on Information Theory

Volume IT-4 December, 1958 Number 4

TABLE OF CONTENTS



Thomas P. Cheatham, Jr.

Thomas P. Cheatham, Jr. (A'52—SM'54) was born in Washington, D.C., on January 30, 1923. He received the B.S. degree from the United States Coast Guard Academy in 1943, and the M.S. and Sc.D. degrees from the Massachusetts Institute of Technology in 1947 and 1952, respectively.

From 1943 to 1946, Dr. Cheatham served as both line and engineering officer in the U. S. Coast Guard, and first became interested in the field of electronics through his close association with the Coast Guard's operational and development program in radar, radar beacons, and Loran.

From 1946 to 1949, he was a research associate at the M.I.T. Research Laboratory of Electronics, where he did work on impulse noise in FM receivers and developed an early interest in statistical communication theory.

In 1949, Dr. Cheatham visited Norway as a consultant to the Norwegian Defense Department, working particularly at Bergen in the formation of a new radar division and assisting in the development of a microwave communication relay system connecting Bergen with Oslo. He returned to the United States in 1950 to complete his graduate studies and to join the Physical Research Laboratories of Boston University as electronic section head and instructor in the Physics Department. During that time, he was principally concerned with research and development on single line scan television systems and in the synthesis of optical and electro-optical devices. In 1952, Dr. Cheatham accepted an appointment as Research Fellow at Harvard University where he carried out research in the field of random processes and statistical communication theory. During the period from 1949 to 1953, he did extensive consultive work in industry and for the government in the field of statistical communications, and particularly on optimal techniques for data processing, criteria for design of prediction computers, network analysis and synthesis.

In 1953, he joined industry on a full-time basis, becoming Director of Research for Melpar, Inc., Boston, Mass. He has directed the growth of this department from its initial formation as a research group to its present size of two laboratories encompassing the broad fields of electronics, physics, and data processing.

Dr. Cheatham is a member of Sigma Xi, and the author of several papers in the field of information theory. He has long been active in both local and national IRE activities. He was elected Chairman of the Boston Section in 1955 and has held various other offices both before and since then. He has been a member of the Information Theory and Modulation System Committee since 1952 and a member of the Administrative Committee since 1956. He was appointed Business Manager in 1957, and elected Chairman of the Professional Group on Information Theory in 1958.

A Broader Base for the PGIT

THOMAS P. CHEATHAM, JR.

It is recognized that the PGIT has achieved at least two of its essential objectives, those of technical stature and a unique identity within its professional society. The evolution of our editorial policies and the high standards that have been achieved for publication of our TRANSACTIONS when combined with the several outstanding symposia that have been sponsored by the PGIT give us a great deal of which to be proud.

However, the achievement of these objectives has been the work of a relatively few individuals whom we might refer to as the "hard core" of our organization. The ability to sustain the present position requires the establishment of an organizational structure that enlarges the amount of active participation from the PGIT membership and which, at the same time, is re-oriented to be compatible with long-range plans and objectives. We feel that our primary purpose is associated with the word "education" and more specifically, the education of the PGIT membership in tune with current interests and trends. Although the PGIT needs reasonably precise and unambiguous constraints to guide it, these should be dynamic in nature rather than static. For example, I feel that the name of our Professional Group on Information Theory is too restricted, and might well be supplemented by an appropriate subtitle to augment and explain the more general and present scope of interest of the whole membership. Such a subtitle might be: "The Transmission and Processing of Information." Thus, the systems engineers in various fields and the computer designer, who apply the basic concepts of information theory to their problems as guide lines in design and as a means of establishing criteria for efficiency and reliability, should have a more obvious position and interest in our group. It is important to recognize the value of this industrial interest in systems and their examination for compatibility with the basic tenets of information theory as a means for providing a logical servo link by which new problems are returned to the basic and applied theoretist.

I would like also to suggest that there is fair justification for recognizing a need for two types of symposia. One is slanted towards the basic research work in the field and has as its principal audience, the serious full-time basic research members of the group. The model for this type of symposia is the M.I.T. symposia of past years that have been held biannually. It is strongly felt that the technical character of this meeting should not change other than to perhaps grow with increasing emphasis on its national and international character. It is suggested that interleaved with this type of symposia and operating on alternate years with it, there should be a broader based type of symposium which emphasizes system applications and fringe areas where we overlap with other professional interest and disciplines. For example, plans are currently in progress for holding a joint Information Theory-Circuit Theory symposium in Los Angeles in June of next year.

To achieve a broader base and a more flexible attitude toward current interest and trends of our membership, it will be necessary to increase the number of working committees and subcommittees so that new interests, new concepts, and new blood can be heard. Special issues of the TRANSACTIONS will be sparked and guided by our newly formed Editorial Board. My hope is that we as a professional group will explore with some spirit of adventure, that we will be organized efficiently, but with sufficient flexibility for enthusiasm and creative thinking to permeate our efforts.

A Statement of Editorial Policy

The IRE TRANSACTIONS ON INFORMATION THEORY is a quarterly journal devoted to the publication of papers on the transmission, processing, and utilization of information. The exact subject matter of acceptable papers is intentionally, by editorial policy, not sharply delimited. Rather, it is hoped that as the focus of research activity changes, a flexible policy will permit the TRANSACTIONS to follow suit and that it will continue to serve its readers with timely articles on the fundamental nature of the communication process. Topics of current appropriateness include extensions of the information theories of Shannon and Wiener and their ramifications, analyses and design of communication systems, information sources, pattern recognition, receiving and detection, automata and learning, large-scale information processing systems, and so forth.

Papers can be of two kinds: tutorial or research, and should be so indicated. The former must be well-written expositions summarizing the state of a field in which research is still in progress, or else bring together as a unity results scattered in the literature. Research papers must be original contributions not published elsewhere. They must present new methods, concepts or ideas, or extend old ones to new areas of applicability; or, they must present new data, findings or inventions, or solve new problems of more than casual interest. They will not be accepted if, in the view of the reviewers and editors, they constitute a straightforward and easy application of existing theory to a special case of limited interest. It is not necessary that the length of each research paper be great; on the contrary, the submission of short but formal research notes is to be encouraged. These will not be published as correspondence, but will be subject to the same review standards as longer papers.

In addition to papers, readers are invited to submit notes to the Correspondence section. These may include early summaries of important work to be published later at greater length, remarks on material that has already appeared, and so forth. Reasonable contributions to this section will be published without editorial review.

All manuscripts should be prepared with clarity of style and economy of mathematical notation always in mind. Unusual symbols are to be avoided and display formulas are to be kept to a minimum consistent with clear exposition. Related work should be adequately referenced and references to readily available literature should be made in place of repeating existing derivations and arguments.

—The Administrative Committee

Time Statistics of Noise*

WILLIAM M. BROWN†

Summary—The measurements made on a system containing noise are usually time averages of the signals, or of quantities defined in terms of the signals. Such measurements are called *time statistics*. The object of this paper is to develop the theory of time statistics and in turn to give methods for calculating them. For the most part the time statistics are formulated in terms of ensemble statistics which are usually provided by statistical mechanics.

If a process consists of, say, all physically realizable models of a system containing noisy resistors, there is no practical way to identify which model one has available for "testing." Thus, a time statistic measured with the available model will not be predictable unless this statistic is the same for almost all the models; when this is the case, the process is called *uniform*[1] *for this statistic*. A dual property is in common use for ensemble statistics. The process is called *stationary for an ensemble statistic*, provided it is the same at all times. Though some discussion of stationarity is given in this paper, the emphasis is on not requiring stationarity. In particular, special attention is given to nonstationarity introduced by determinate signals. While stationarity plays only a minor role in the theory of the time statistics of noise, uniformity plays a crucial role. Given only uniformity, Theorem 1 formulates time statistics as the time average of the corresponding ensemble statistics. The additional condition of stationarity merely simplifies the calculation by rendering the "ergodic hypothesis" satisfied, *i.e.*, by rendering equality of time and ensemble statistics.

With Theorem 1 as a nucleus, the remainder of the paper attempts to develop an understanding of what makes a process uniform. There is no attempt to give detailed proofs, but there is an effort to maintain a clear distinction between physical motivations, the definitions, and the theorems. Some elementary sample calculations of practical interest are included; these serve to illustrate several parts of the theory. Though calculations involving such problems as the evaluation of difficult integrals do arise in some applications of the theory, simple samples have been used here, since they are adequate as an aid to understanding the theory.

I. Introduction

BACKGROUND material on noise theory is found in the Bibliography. Lucid accounts of the history of ergodic theory are given [6], [14]; Loeve [10] briefly brings the history up to date. Practical insight for the work is provided in a very helpful manner [1], [2], [8], [9], [11], as is the mathematical background [1], [3], [4], [7], [10], [12].

Motivation for this paper results from the inadequate descriptions in the existing literature of the ties between probabilistic notions associated with noise ensembles and time-average measurements. In particular, it is intended that this paper fill the gap between brief discussions on ergodicity of the type found in the engineering literature (cf. [8], p. 114) and mathematical details such as those given by Doob [4] and Loeve [10]. Section II of this paper gives a fresh organization to the theory underlying time statistics. Though the technique in the conclusion of Theorem 1 has been used by other workers, the theorem has not been publicized nor has it been used elsewhere to organize the theory.

While some of the material presented here is new, the exact status of the various parts is not described; the general organization and Section II-D are probably the major contributions.

The following characterize noise analysis. A collection of physical things (models), rather than only one thing, must be analyzed, and average values are the quantities of chief concern. The collection is a physical necessity when one wishes to analyze a system for which certain factors are not completely specified. It is common not to specify certain details about the microscopic structure of elements, such as noisy resistors. If only macroscopic properties, such as resistance value, are fixed, different realizations give rise to different noise waveforms. Analysis, which is to be applicable to any realization, is limited in that usually only average values are predictable.

Stated briefly, a *noise process* is a (finite) set of real-valued functions of two variables. Each function is called a noise ensemble. The first variable (typical point denoted by alpha) depicts the member of the ensemble (also called model or realization here); its domain is assumed to be a probability measure space I. The second variable is usually over the real line[2] (time).

Fig. 1 illustrates the models of a noise process. For example, each model could be a particular realization of a certain type of radar and the ensemble could consist of all possible realizations. The functions would be various electrical quantities. The f-noise ensemble, which is composed of $f_\alpha(t)$, $f_\sigma(t)$, $f_\beta(t)$, $\cdots$ would be formed by taking these time functions from corresponding places, such as the input voltages of the α radar, σ radar, β radar, $\cdots$, respectively; the g ensemble, which is composed of g_β, g_α, g_σ, $\cdots$ might be taken as the currents from some particular wire in the respective radars, etc. It may be that almost all functions of time in, say, the f ensemble are the same, in which case the f ensemble is called *determinate*. Typically, the intended signals in a system are determinate, save possibly for absolute time reference,

* Manuscript received by the PGIT, April 25, 1958. This paper is based on the author's Doctoral dissertation. Some of the work was supported by USAF Contract No. 33(616)-3374 at the Radiation Lab., The Johns Hopkins University, Baltimore, Md.

† Elec. Eng. Dept., University of Mich., Ann Arbor, Mich.

[1] The term "ergodic" is reserved for a condition that makes a process uniform for a large class of statistics. Also, it should be mentioned that "almost all" means for all models, except possibly a set of measure zero, and that terms in italic are being defined.

[2] The domain for the second variable might be only the integers, in which case the process is called *discrete*. Actually, it is easy to adjust this paper to cover any set for the second variable, provided there is a notion of average (denoted by A later) and a notion of translation associated with this variable. The author takes the real line only to be specific.

but this is accounted for in the section on determinate signals (II-D).

If F is a real valued function of α, the ensemble average (expected value) of F is denoted by $E(F)$ or by $\int_I F(\alpha)du(\alpha)$. A lengthy discussion of this *integral* is not included, but the following basic properties are mentioned. E is a linear operator; if $F(\alpha) = 1$ for each α in I, $E(F) = 1$; and if $F(\alpha) \geq 0$ for each α in I and $E(F)$ exists, $E(F) \geq 0$. Also, if B is a set of points in I and $\phi_B(\alpha)$ is taken as one when α is in B, and as zero when α is not in B, then B is measurable if, and only if, $E(\phi_B)$ exists and $E(\phi_B)$ is the *measure of* B. Finally, if the set of α's for which $F(\alpha) \geq C$ is measurable for each real C, then F is called a *measureable* function of α.

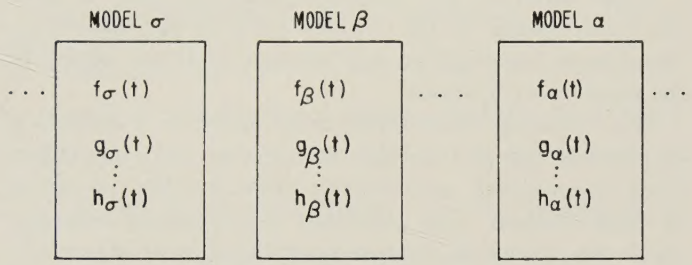


Fig. 1—Ensemble of models.

Similarly, if F is a function of time, the time *average of* F is denoted by $A(F)$ and defined by

$$2A(F) \equiv \lim_{a\to\infty} \frac{1}{a} \int_0^a F(t)\, dt + \lim_{b\to\infty} \frac{1}{b} \int_{-b}^0 F(t)\, dt.$$

("$\equiv$" means defined as.)

Taking either of these terms alone also provides a useful definition of the time average of F, and using limits in the mean (with respect to α integration, where F is assumed to be a function of α as well as t) also provides useful definitions of time average. We take the indicated definition in order to be specific. Observe that if $A(F)$ exists,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^T F(t)\, dt$$

exists; this is sometimes used for the definition of A. However, the definition used here has the advantage that the existence of $A(F)$ is all that is needed to prove that $A(F)$ is independent of phase, *i.e.*, $A[F(t)] = A[F(t + c)]$ for all c.

Now let F be a function of α and t. The *ensemble statistic* for F is defined as $E(F)$. The process is *stationary for* F if $E(F)$ is the same for all t. The *time statistic* for F is defined as $A(F)$ and the process is *uniform for* F if $A(F)$ is the same for almost all α. "Almost all α" will sometimes be abbreviated by a.e. (almost everywhere). Before proceeding with the theory, some examples of typical relations between F and the noise ensemble will be given.

The *nth cumulative distribution* of f. If $J(y_1, \cdots, y_n) \equiv 1$ when every $y_k \geq 0$ and $J \equiv 0$ otherwise, then the nth ensemble distribution of f evaluated at $(x_1, x_2, \cdots, x_n; t_1 + t, t_2 + t, \cdots, t_n + t)$ is given by $EJ[x_1 - f_\alpha(t_1 + t), \cdots, x_n - f_\alpha(t_n + t)]$. If E is replaced by A, one has the nth time distribution of f. Of course, if the x's and t's are fixed, J above can be considered a function F of α and t.

A *joint characteristic function* of f and g. Here one may take

$$F(\alpha, t) = \exp\left[i \sum_{k=1}^{n} f_\alpha(t_k + t)q_k + i \sum_{j=1}^{m} g_\alpha(u_j + t)r_j \right].$$

Then the joint ensemble characteristic function of f and g, evaluated at the t_k, q_k, u_j, and r_j values indicated, is given by $E(F)$, and $A(F)$ gives the corresponding time statistic.

If P represents all the parameters, such as the q and r variables in the last example,

$$f(\alpha, t, n) \equiv [f_\alpha(t_1 + t), f_\alpha(t_2 + t), \cdots, f_\alpha(t_n + t)],$$

$$g(\alpha, t, m) \equiv [g_\alpha(u_1 + t), \cdots, g_\alpha(u_m + t)],$$

etc., then the general (joint) statistic for the ensembles $f, g, \cdots, h$ is formulated as follows. Let

$$F(\alpha, t) = H[f(\alpha, t, n), g(\alpha, t, m), \cdots, h(\alpha, t, p); P].$$

The ensemble statistic for this F is $E(F)$ and the time statistic is $A(F)$. Here H denotes a function of many real variables; the values of H (and hence of F) are either real or in a cartesian space having a dimensionality greater than one. In the latter case, H can be considered an ordered n-tuple of real-valued functions and the average (E or A) of H is taken as the ordered n-tuple of averages. One might take H to be a more general mapping than that provided by a function of many real variables, but again it seems better to be specific than to strive for more generality.

II. Fundamental Theorems

The program is first to give Theorem 1 which, though easy to prove, is important, comprehensive, and useful. The remainder of the section supplements Theorem 1 largely by describing conditions that render a process uniform. The development might be compared to the theory of Fourier series in that one can first display the basic nature and use of Fourier series by assuming that certain manipulations are permitted and then proceed to acquire insight by discovering what makes the manipulations valid. Roughly, this section (after Theorem 1) displays the following: how indecomposability (metric transitivity) implies uniformity for a specific F, how this is generalized to a large class of statistics for nondeterminate signals—with stationarity then brought in to render the ergodic hypothesis satisfies, and finally how ergodicity joined with independence over time implies uniformity when determinate and nondeterminate signals are present.

A. Time Average of Ensemble Statistics

As pointed out in the Introduction, essentially all statistics can be formulated as averages of a function $F(\alpha, t)$. In applications of the following theorem, $E(F)$ is usually given and then $A(F)$ is to be calculated.

Theorem 1: Let the process be uniform for F(*i.e.*, let $A(F)$ be the same a.e.) and let $AE(F) = EA(F)$; then $A(F) = AE(F)$ for almost all members. (The converse is also true.)

To see this we start with $AE(F) = EA(F)$; but $A(F)$ is constant a.e. and hence $EA(F)$ is equal to this common value. That is, $A(F) = AE(F)$ a.e.

In the hypothesis of this theorem there are three items for concern, *i.e.*, the existence of $A(F)$, the interchange of the intergrations defining A and E, and the uniformity of the process. Theorems that give the existence of $A(F)$ are called ergodic theorems and these receive much attention in the mathematical literature; the two most famous ergodic theorems are stated later. Standard integration theory can be used to justify the interchange of A and E—the following lemma being typical. Finally, uniformity is given rather thorough study in this paper.

Lemma: Let F be (*Lebesgue-Stieltjes*) measurable on the cartesian product space of I and the real line where the usual product measure is used; let F be bounded, and let $A(F)$ exist almost everywhere on I; then $AE(F)$ and $EA(F)$ exist and are equal. Standard theorems, such as Fubini's on interchanging the order of integrations, are the major tools required for the proof of this lemma. The "ergodic hypothesis" (cf. [8], p. 114 or cf. [6], p. 52) will now be derived as a corollary to Theorem 1.

Corollary: Let the process be stationary as well as uniform for F; then, granting $AE(F) = EA(F)$, $A(F) = E(F)$.

From the theorem, $A(F) = AE(F)$, but by stationarity $E(F)$ is independent of time and hence, $AE(F) = E(F)$.

A sample calculation using Theorem 1 will be discussed. This example requires the generality of Theorem 1; however, the two calculations on modulated waves given in Section III might be handled by the more restricted results given in Section II-D. Here we consider a sampling circuit that samples the input f at times $\{t_k\}$. The output is denoted by h and the sampling functions, which are determinate, are denoted by $\{g_k\}$. Then the equation giving the output in terms of the input is assumed to be

$$h_\alpha(t) = \sum_{k=-\infty}^{\infty} f_\alpha(t_k) g_k(t).$$

First let us observe that from the linearity of this equation it can easily be shown that if all of the ensemble distributions of f are Gaussian, all the ensemble distributions of h will be Gaussian (cf. [2]). However, even if f is stationary, h will not be stationary (except for first-order statistics as shown below).

Next, consider the special case that one has if $g_k(t) = 1$ for $t_k \leq t < t_{k+1}$ and $g_k(t) = 0$ for other values of t. Under this condition it may be shown that the first time distribution of h is equal to the first ensemble distribution of f. For this it is assumed that f is stationary for its first ensemble distribution and that Theorem 1 applies. Let P_f and P_h denote the first ensemble distributions of f and h, respectively. Now, when $t_k \leq t < t_{k+1}$, $h(t) = f(t_k)$ and hence $P_h(x, t) = P_f(x, t_k)$. Since f is assumed stationary, $P_h(x, t) = P_f(x)$. Taking the time average is effortless and if $\bar{P}_h(x)$ denotes the time distribution of h, Theorem 1 gives $\bar{P}_h(x) = P_f(x)$ which was to be shown.[3] If the sampling points are far apart and these flat g_k are used, h must have its power concentrated at low frequencies even though f may have its power concentrated at high (more easily generated) frequencies. Thus this sampler can be viewed as a spectrum compressor; for additional discussion of this point see [17]. A sampler was built for this case and the above result, $\bar{P}_h = P_f$, was checked using a Gaussian f and a special low-frequency distribution analyzer. The agreement was good.

Finally, to make the calculation look less trivial, let us consider another form for the g_k. The value $f(t_k)$ is usually held on a capacitor; however, there is some discharging between samples. This can be accounted for by taking $g_k(t) = \exp[(t_k - t)/\tau]$ for $t_k \leq t < t_{k+1}$. For said t, $h_\alpha(t) \leq x$ if $f_\alpha(t_k) \leq x \exp[(t - t_k)/\tau]$. Thus, for $t_k \leq t < t_{k+1}$, we have $P_h(x, t) = P_f x[e^{(t-t_k)/\tau}, t_k]$. Now, let f be stationary for P_f and let $t_{k+1} - t_k = T$ for each k. Then $\bar{P}_h(x) = AP_h(x, t)$ leads to $\bar{P}_h(x) = 1/T \int_0^T P_f(x\, e^{t/\tau})dt$. In this case, even if P_f is Gaussian, $\bar{P}_h$ will not be Gaussian. To look at this result for a Gaussian f it is more convenient to consider density distributions. Let $\bar{W}_h$ and W_f denote the density distributions for h and f, *e.g.*, $W_f = \partial/\partial x\, P_f$. Then assuming that differentiation under the integral sign is permitted, we get

$$\bar{W}_h(x) = \frac{1}{T} \int_0^T W_f(xe^{t/\tau}) e^{t/\tau}\, dt.$$

When W_f is a Gaussian density distribution having zero mean value, it is easy to show that this gives

$$\bar{W}_h(x) = (\tau/Tx)[\operatorname{erf}\left(\frac{x}{\sigma} e^{T/\tau}\right) - \operatorname{erf}(x/\sigma)]$$

where

$$\operatorname{erf}(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{z} \exp(-\theta^2/2)\, d\theta.$$

Another point that could be made here is that the equation for $\bar{W}_h$ given above is equivalent to $\bar{W}_h(x) = AW_h(x, t)$; hence, with the required interchange of differentiation and integration, the conclusion of Theorem 1 has been extended to a "statistical quantiy" other than those falling under the formulations in the Introduction. This point is given additional discussion in Section III-A.

B. Indecomposability Implies Uniformity

We first make some preliminary definitions. With α and τ fixed, $F(\alpha, t + \tau)$ is a function of t; a point θ in I is a *τ translate of α for F* if $F(\theta, t) = F(\alpha, t + \tau)$ for all t (if any such θ exists). The trajectory of α is the set of all translates found for all real τ. In other words, a trajectory is a set of points in I such that the corresponding functions

[3] The relation is not so simple for higher-order distributions.

of t furnished by F are the same, except for phase. A process is *indecomposable for* F if measurable collections of trajectories for F always have either measure zero or one. Before giving an example of a (nontrivial) indecomposable space, the important theorem is given.

Theorem 2: If the process is indecomposable for F and $A(F)$ is measurable, the process is uniform for F.

To prove this, we first note that $A(F)$ is the same for every point on any particular trajectory,[4] *i.e.*, the time average of a function is independent of its phase. Thus, the set of points in I having $A(F) \leq C$ is a set of trajectories. Indecomposability with the measurability of $A(F)$ implies that this set is of measure zero or one, depending on C. This implies that $A(F)$ as a function of α is constant a.e., *i.e.*, the process is uniform for F.

To get some insight as to the nature of indecomposable spaces, an example will be given. If the individual trajectories are measurable, a space I either has a single trajectory of measure one or an (uncountable) infinity of trajectories, each of measure zero. The single trajectory case might be considered trivial, though sometimes of interest. Illustrating a nontrivial example is important since it shows that one may have indecomposability (even) when the models in the ensemble differ by more than merely time translations. A. Novikoff suggested the following example. The space I is taken as the unit square in the (x, y) plane where it is assumed that there is a model corresponding to each $\alpha = (x, y)$ in the square. Fig. 2 depicts the space with part of a typical trajectory

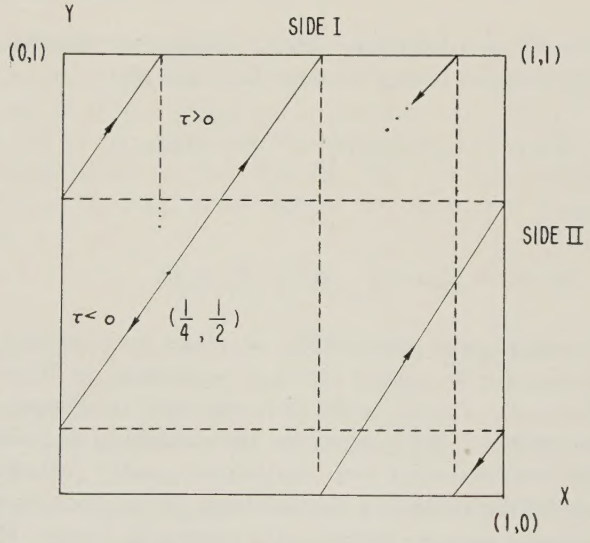


Fig. 2—Example of indecomposable space.

drawn in. Let α be a fixed point in the square, say $(\frac{1}{4}, \frac{1}{2})$; the τ translate of α is assumed to be on a line of irrational slope, say $\sqrt{2}$, at a distance equal to τ along the indicated trajectory where positive τ rides upward. If the distance τ causes "puncture" of side I, the trajectory returns on the x axis at the puncture abcissa. Starting with another point not on this trajectory, say $(\frac{1}{4}, \frac{1}{4})$, another trajectory is formed using the same slope, etc. For this set of trajectories it can be shown that any collection of trajectories will be of Lebesgue measure one, zero, or nonmeasurable. An F that induces the above set of trajectories is the following, where $\alpha = (x, y)$:

$$F(\alpha, t) = \sin 2\pi(x + t/\sqrt{3}) + \sin 2\pi(y + \sqrt{2}t/\sqrt{3}).$$

C. Ergodicity

Normally one wants a process to be uniform (and/or stationary) for a large class of statistics, *i.e.*, for a large class of functions H and for all values of the t_k's, u_k's, $\cdots$, and P. Strong forms of indecomposability and stationarity will bring this about.

One may consider the function of α and t formed by taking the ordered n tuple of nondeterminate noise ensemble. Thus,

$$K(\alpha, t) \equiv [f_\alpha(t), g_\alpha(t), \cdots, h_\alpha(t)].$$

A process is *basically indecomposable* if it is indecomposable for K. For the remainder of this paper it is convenient to think of the process as consisting of a set of nondeterminate ensembles, such as resistor noise sources, and a set of determinate ensembles, such as the intended signals. Then, in the definition of K, only the nondeterminate signals are included. The situation presented by the mixture is developed in the next section.

Theorem 3: If the process is basically indecomposable, it is indecomposable for any H. In turn, in accordance with Theorem 2, the process will be uniform for any H.

To prove this, one may observe that a τ translate of α for K is necessarily a τ translate for H, that is, if

$$[f_\alpha(t + \tau), \cdots, h_\alpha(t + \tau)] = [f_\theta(t), \cdots, h_\theta(t)]$$

for all t, with α, θ, and τ fixed, then

$$H[f(\alpha, t + \tau, n), \cdots, h(\alpha, t + \tau, p); P] = H[f(\theta, t, n), \cdots, h(\theta, t, p); P] \quad \text{for all} \quad t.$$

From this it follows that each trajectory for H will consist of a collection of sets of trajectories for K. In turn, the indecomposability for K will imply indecomposability for H.

We now want to define a strong form of stationarity. If B is a set of points in I, let B_τ denote the set of all θ's in I such that there exists an α in B for which θ is a τ translate of α for K. The process is *strictly stationary* (a weaker definition is sometimes used, cf. [4]) if for any measurable B_0, $B\tau$ is measurable and their measures are equal for all τ. If the F given with Fig. 2 is equal to K (*i.e.*, there is just this one noise in the process), the example given there is a strictly stationary process.

Theorem 4: If the process is strictly stationary and $E(H)$ exists, $E(H)$ is independent of t, *i.e.*, the process is stationary for H.

Because of the assumed stationarity, the values of H

[4] This theorem also holds for the other useful definitions mentioned for $A(F)$ in the Introduction.

merely move along trajectories as "t" varies, and the amount of the measure space covered by each value interval of H remains constant. From this consideration it is not difficult to see that $E(H)$ will be independent of time.

A process is *ergodic* if it is basically indecomposable and strictly stationary. Combining Theorems 1, 3, and 4 we have the following.

Theorem 5: Let H be any function over the nondeterminate ensembles, let $E(H)$ and $A(H)$ exist with $A(H)$ measurable, and let the process be ergodic; then $A(H) = E(H)$ a.e.

One can observe that the theory developed here would be rather empty if time statistics seldom existed. Ergodic theorems give conditions under which these time averages exist in some sense. Only the two most famous ergodic theorems (cf. [10], p. 410) are stated. First assume that the process is strictly stationary. Also for each α in I and each τ it is assumed that there is one, and only one, τ translate of α for K. Birkhoff's ergodic theorem then concludes that the additional condition that $E(H)$ exists implies that $A(H)$ exists a.e. (and is finite). Von Neumann's ergodic theorem concludes that if $E(H^2)$ exists, l.i.m.$_{T\to\infty}$ $1/2T \int_T^T H\, dt$ exists. These theorems are quite deep and the proofs are rather long, requiring many results from integration theory.

D. Independence Over Time

By considering examples, it can be shown that a process may not be indecomposable for joint statistics of determinate and nondeterminate ensembles. This can easily be the case even if the ensemble is enlarged in a reasonable way to include all translations of the originally determinate signals. Thus, the crucial property of uniformity must result from some other condition. Some examples display the fact that independence over time is this condition.

To simplify this section it will be assumed that there is only one determinate ensemble $g(t)$ and only one nondeterminate ensemble $f_\alpha(t)$. It is easy to generalize here to include many ensembles in these classes, but the equations become somewhat lengthy. The problem to be considered here is that of determining joint time statistics of f and g, *i.e.*, one has a function, $H[f(\alpha, t, n)\ g(t, m);\ P]$, and desires to show that the process is uniform for H. Then Theorem 1 can be used to give $A(H) = AE(H)$. To simplify the notation, the n, m, and P in H are not written in this section. If h is a function of f and g (as in the examples of Section III), statistics of h are included as special cases of joint statistics of f and g. Finally, it is convenient to append a subscript to the time average operator to indicate which variable is being averaged over, *e.g.*,

$$A_s H[f(\alpha, s),\ g(t)] \equiv \lim_{a\to\infty} \frac{1}{2a} \int_0^a H[f(\alpha, s),\ g(t)]\, ds + \lim_{b\to\infty} \frac{1}{2b} \int_{-b}^0 H[f(\alpha, s),\ g(t)]\, ds.$$

Now let α be a particular member of the ensemble, then f_α *and* g *are independent over time* for H if

$$A_t H[f(\alpha, t),\ g(t)] = A_s A_r H[f(\alpha, s),\ g(r)] = A_r A_s H[f(\alpha, s),\ g(r)].$$

A similar definition of independence over the ensemble can be made; however, this type of independence is not of interest here and hence independence means the above type in this paper.

It is important physically that the time statistic not be sensitive to the relative phase of f_α and g. Investigating this leads to Theorem 6, which is an alternative definition of independence. Also, it should be noted that the above definition of independence is essentially equivalent to the more conventional definition given as the equality of joint distributions and the product of individual distributions.

Theorem 6: $A(H)$ is independent of the phase of g if $f_\alpha(t)$ and $g(t + c)$ are independent for all c. It is assumed here that certain interchanges[5] of operators are permitted.

The formal proof is easy and is omitted. An interesting point about this theorem is that there is no analogous theorem when the real line domain for t is replaced with an arbitrary measure space not having a notion of translation associated with it.

Theorem 7: Let the process be ergodic (*i.e.*, let the process be strictly stationary and indecomposable for f), then granting some interchanges the process is uniform for H if, and only if, f_α and g are independent for H for almost all α.

To prove the "if" part of this theorem we have an ergodic process with f and g independent for H. Thus,

$$A(H) = A_s A_r H[f(\alpha, r),\ g(s)].$$

Here $A_r(H)$ is a time statistic of f, since the values of g are fixed during the r integration, and hence, by ergodicity this time average is equal to $E(H)$ a.e. This gives

$$A(H) = A_s E H[f(\alpha, r),\ g(s)],$$

which says (among other things) that $A(H)$ is the same for almost all α, which was to be shown.

It might be observed that the $E(H)$ part is independent of r by the assumed stationarity.

For the "only if" part of the theorem, we have an ergodic process with $A(H)$ contant a.e. By Theorem 1

$$A(H) = AEH[f(\alpha, t),\ g(t)].$$

During the expected value integration, the g values are fixed and hence $E(H)$ can be considered an ensemble statistic of f. Then by ergodicity, $E(H)$ can be replaced by the average over the t appearing in H. Thus, we get

$$A(H) = A_t A_s H[f(\alpha, s),\ g(t)].$$

[5] Lemmas, such as the one given with Theorem 1, can be used to justify the interchanges; however, these add much length and little content to the points of interest here.

Granting a final interchange f_α and g are independent for almost all α which completes the proof.

The above theorem displays the importance of independence and the next theorem adds some insight.

Theorem 8: Let the process indecomposable for f and let f_α and g be independent for H a.e., then the process is uniform for H—assuming AH is measurable.

To prove this it is first shown that $A(H)$ is constant on a trajectory. Let θ be a τ translate of α. Then AH at θ is given thus (by independence and the notion of τ translate):

$$A_\tau A_s H[f(\theta, s), g(r)] = A_\tau A_s H[f(\alpha, s + \tau), g(r)].$$

But it is easy to see that $A_s(H)$ here is independent of τ, *i.e.*, time averages are not sensitive to phase. Hence, this last form is equivalent to

$$A_\tau A_s H[f(\alpha, s), g(r)] = A(H)$$

at α, which was to be shown. Now as was done with Theorem 2, the set of α's for which $A(H) \leq C$ is a collection of trajectories for f. By the assumed indecomposability this set is of measure zero or one depending on C, *i.e.*, $A(H)$ is constant a.e. which completes the proof.

III. Sample Calculations

This section contains calculations of the first time distribution of a sine wave amplitude modulated by Gaussian noise and of a periodic wave angularly modulated by "almost any" signal. For simplicity these calculations have been restricted to first-ordered statistics; of course, Section II is applicable to any statistic. Experimental verifications are included.

A. Amplitude Modulated Sine Wave

Here we take $h_\alpha(t) = f_\alpha(t) \sin wt$. Since h is a linear transformation of f, h will have Gaussian ensemble statistics when f is Gaussian. However, it is shown in this section that the nonstationarity of h renders its time statistics strikingly non-Gaussian. Let W_f be the first ensemble density distribution of f and let P_f be the corresponding ensemble cumulative distribution.[6] Of course, $W_f(x, t) = \partial/\partial x \; P_f(x, t)$. Let W_h and P_h be the corresponding functions for h. If $\sin wt \neq 0$ we have $h_\alpha(t)$ in the interval $(z, z + dz)$ if $f_\alpha(t)$ is in the interval $(z/\sin wt, z/\sin wt + dz/\sin wt)$. From this it follows that $W_h(z, t) = W_f(z/\sin wt, t) \mid \sin wt \mid^{-1}$. If f is stationary for W_f and we set $\bar{W}_h(z) = A\, W_h(z, t)$, the above equation gives

$$\bar{W}_h(z) = \frac{1}{2\pi} \int_0^{2\pi} W_f(z/\sin \theta) \mid \sin \theta \mid^{-1} d\theta.$$

Before evaluating this for a Gaussian W_f, let it be noted that there is a problem posed here. Density distributions are not statistics of the form H given in the Introduction, but cumulative distributions are such statistics. As mentioned in the example at the end of Section II-A, the technique in the conclusion of Theorem 1 can be extended to density distributions if the appropriate integration and differentiation can be interchanged. This may be improvised as needed. In particular, it is justifiable in the case here and in the example of Section II-A under the single assumption that P_f is absolutely continuous. Continuing in this manner, many important extensions can be formulated. For example, one may formulate the power density spectrum over time of a nonstationary noise f as the time average of the Fourier transform of the ensemble correlation. Thus,

$$A_t \int_{-\infty}^{\infty} \exp(-jw\tau) E[f_\alpha(t + \tau) f_\alpha(t)]\, d\tau.$$

Here we interchange the Fourier and A_t operators.

Completing the above calculation, when W_f is Gaussian having zero average value, gives

$$\bar{W}_h(z) = \frac{2}{\pi} \int_0^{\pi/2} (\sqrt{2\pi}\sigma)^{-1} \exp(-z^2/2\sigma^2 \sin^2 \theta) \csc \theta \, d\theta.$$

Letting $y = (\cot \theta)^{1/2}$, one gets the form given in Grobner and Hofreiter [5], p. 57. The result is

$$\bar{W}_h(z) = (\pi \sqrt{2\pi}\sigma)^{-1} \exp(-z^2/4\sigma^2) K_0(z^2/4\sigma^2).$$

Fig. 3 gives this density distribution along with a Gaussian distribution having the same standard deviation. Fig. 4 shows these same data taken experimentally. The multiplication was done by mixing a sine wave at x band and a low frequency noise in a magic Tee. The signal was then beaten down to a 30-mc carrier and analyzed on an analyzer similar to the one described in [13]. Fig. 5 illustrates the quality of the multiplication.

B. Angle Modulated Waves

Let $h_\alpha(t) = F[wt + f_\alpha(t)]$ where F is periodic with period "a" and w is constant. The first probability distribution over time of h is calculated in this part. In fact, it is shown that this distribution is not influenced by a large class of modulating signals. However, the higher order distributions are influenced by the modulation. Since F is periodic, wt can be replaced with a saw-tooth wave $g(t)$ where $g(t) = wt$ for $0 \leq t < a/w$ and g has period a/w.

Drawing on Section II-D, it may be assumed that f_α and g have a certain amount of independence; it is easy to choose f_α dependent on g in ways that cause the distribution of the modulated wave to differ from the distribution of the unmodulated wave. The problem is to calculate the first distribution over time of h_α. If we let $\bar{P}_h$ denote this distribution (assumed to be the same for all α) and let $H(v) = 1$ if $v \geq 0$ and $H(v) = 0$ if $v < 0$, then

$$\bar{P}_h(z) = A_t H\{z - F[g(t) + f_\alpha(t)]\}.$$

[6] It is assumed that P_f is absolutely continuous.

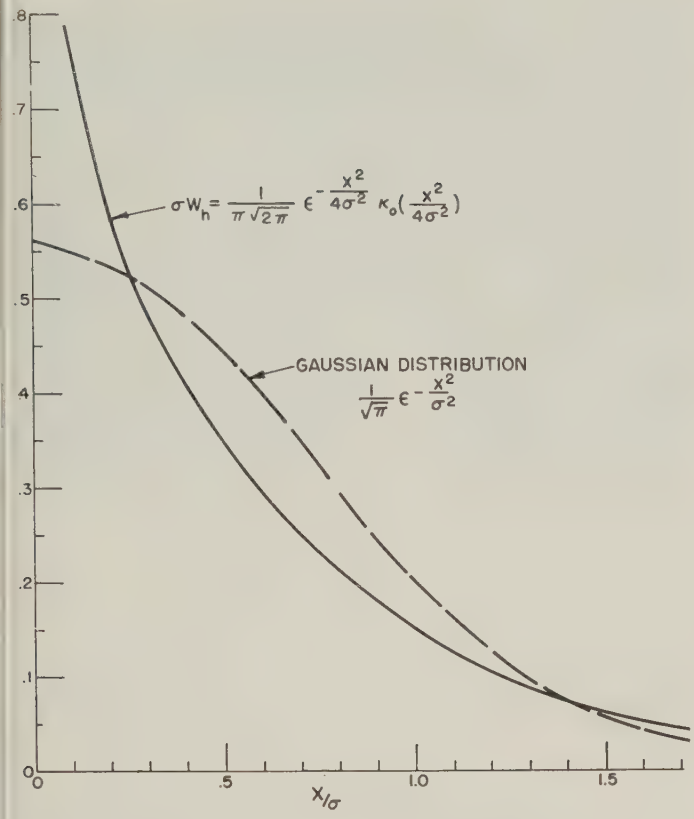


Fig. 3—Density distribution of sine wave amplifier modulated by Gaussian noise.

The first method of doing the calculation is to assume that f_α and g are independent for H, then

$$\bar{P}_h(z) = A_s A_r H\{z - F[g(r) + f_\alpha(s)]\}.$$

However, $A_r H$ is recognized as the first distribution of the phase shifted wave $F[g(t) + \theta]$ where $\theta = f_\alpha(s)$. Clearly, by the periodicity of F, this phase shift does not change the distribution. Thus, if $\bar{P}_{h_0}$ denotes the distribution of the unmodulated wave,

$$\bar{P}_h(z) = A_s \bar{P}_{h_0}(z) = \bar{P}_{h_0}(z).$$

The type of independence assumed above leaves something to be desired in that it depends on F, *i.e.*, one could have the independence for certain F's but not have it for others. It will now be shown that if g and f_α have ordinary first-order independence, $\bar{P}_h(z) = \bar{P}_{h_0}(z)$ for any measurable F. Here it is assumed that the joint distribution of f_α and g over time is equal to the product of the first distributions of f and g. Using the basic theorem on calculating average values from distributions (cf. [4], p. 12) we have

$$\bar{P}_h(z) = A_t H\{z - F[g(t) + f_\alpha(t)]\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H[z - F(x + y)]\, d\bar{P}_{fg}(x, y).$$

The assumed independence gives

$$\bar{P}_h(z) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} H[z - F(x + y)]\, d\bar{P}_g(x)\right] d\bar{P}_f(y).$$

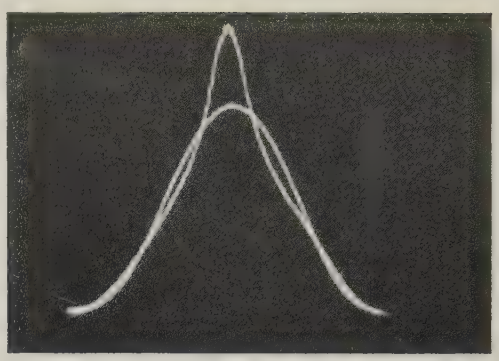
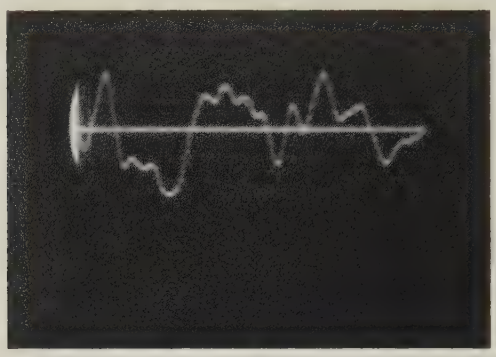
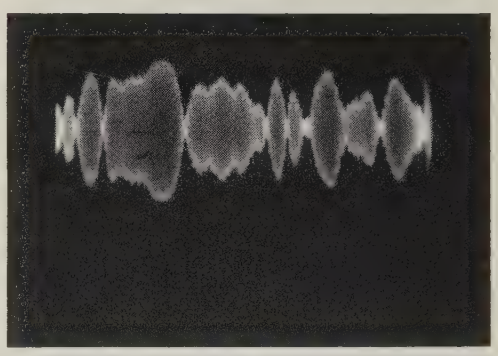


Fig. 4—Density distribution of sine wave amplitude modulated by Gaussian noise. Double exposure. Receiver (Gaussian) noise and mixed noise. Compare Fig. 3.



(a)



(b)

Fig. 5—Synchronized pictures of modulation and modulated signals. (a) Modulating signal. (b) Modulated signal.

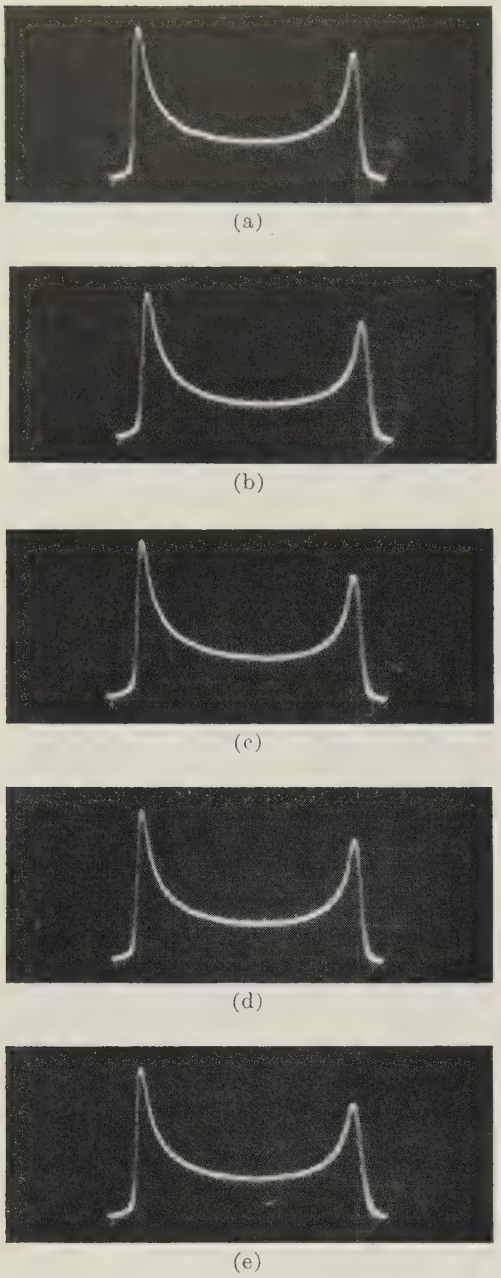
Since $g(t)$ is a saw-tooth,

$$\bar{P}_h(z) = \int_{-\infty}^{\infty} \frac{1}{a}\left[\int_0^a H[z - F(x + y)]\, dx\right] d\bar{P}_f(y).$$

Because of the periodicity of F, the term in brackets is independent of y and we have (observing that the term in brackets is equal to $\bar{P}_{h_0}$)

$$\bar{P}_h = P_{h_0}(z) \int_{-\infty}^{\infty} d\bar{P}_f(y) = \bar{P}_{h_0}(z).$$

This result was experimentally verified by using a sine wave for F and a variety of modulating signals (f_α). The results are shown on Fig. 6. The frequency modulated signal (h) was produced by mixing two klystron outputs. One klystron was at a fixed frequency while the frequency



(a) (b) (c) (d) (e)

Fig. 6—Density distribution of frequency-modulated sine wave. (a) Zero modulation. (b) Frequency modulation with 20-kc block of noise from G. R. noise source. Rectified average of 0.3 volt on repeller. (c) Same as (b), except 0.6 volt on repeller. (d) Noise from G. R. noise source, amplified and clipped, and then put on repeller. Rectifier avarage of 0.33 volt on repeller. (e) Frequency modulation with 300-kc sine wave of 0.5 volt (rms) on repeller. Note: Carrier frequency of 3 mc throughout. Repeller sensitivity of approximately 2.5 mc per volt.

of the other was varied by applying the modulating signal to its repeller. The modulation was broad and the spectral spread was observed on a spectrum analyzer. The equipment is pictured in Fig. 7. The density analyzer is the same as the one used for Fig. 4.

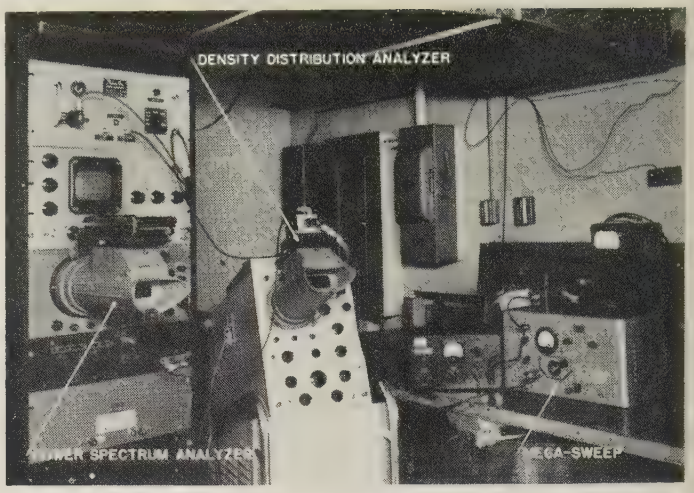


Fig. 7—Apparatus for frequency modulation.

IV. Acknowledgment

The author is very grateful to his many teachers and friends for their help. In connection with this work, he is particularly indebted to Profs. W. C. Gore and W. H. Huggins.

Bibliography

[1] A. Blanc-Lapierre and R. Fortet, "Theorie Des Functions Aleatories," Masson and Company, Paris, France, 1953.
[2] W. M. Brown, "Noise Statistics After Transformations Commonly Found in Circuits" Dept. of Commerce, Office of Tech. Services, Washington, D. C., OTS No. PB121515; 1955.
[3] H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J.; 1951.
[4] J. L. Doob, "Stochastic Processes," John Wiley and Sons, Inc., New York, N. Y.; 1953.
[5] W. Grobner and N. Hofrieter, "Integraltafel," Springer-Verlag Co., Inc., Vienna, Austria; 1950.
[6] A. I. Khinchin, "Statistical Mechanics," Dover Press, New York, N. Y.; 1949.
[7] A. N. Kolmogoroff, "Foundations of the Theory of Probability," Chelsea Publishing Co., Inc., New York, N. Y.; 1950.
[8] J. H. Laning and R. H. Battin, "Random Processes in Automatic Control," McGraw-Hill Book Co., Inc., New York, N. Y.; 1956.
[9] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," McGraw-Hill Book Co., Inc., New York, N. Y.; 1950.
[10] M. Loeve, "Probability Theory," D. Van Nostrand Co., Inc., New York, N. Y.; 1955.
[11] D. Middleton and D. Van Meter, "Detection and extraction of signals in noise from point of view of statistical decision theory," *J. Soc. Ind. Appl. Math.*, vol. 3, pp. 192–253; December, 1955, and vol. 4, pp. 86–119; June, 1956.
[12] M. E. Munroe, "Introduction to Measure and Integration," Addison-Wesley Co., Cambridge, Mass.; 1953.
[13] L. W. Orr, "Wideband Amplitude Distribution Analyzer of Voltage Sources," Electronics Defense Group, University of Michigan, Ann Arbor, Mich., Tech. Rep. No. 22; 1953.
[14] D. ter Haar, "Elements of Statistical Mechanics," Rinehart and Co., Inc., New York, N. Y.; 1954.
[15] N. Wax, ed., "Noise and Stochastic Processes," Dover Publications, Inc., New York, N. Y.; 1949.
[16] N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," M.I.T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, N. Y.; 1949.
[17] D. F. Winter, "A Gaussian noise generator for frequencies down to zero cycle per second," 1954 IRE Convention Record, vol. 2, pt. 4, pp. 23–29.

Multiple Error Correction by Means of Parity Checks*

GERALD E. SACKS†

Summary—An n-place binary parity check code which corrects up to and including e errors in each code letter is fully described by its n characteristics, which are r-dimensional vectors, where r is the number of redundant binits in each code letter. It is shown that the characteristics of such a code have the essential property that any subset of $2e$ of them are linearly independent. An upper bound on r for fixed n and e is obtained by consideration of a systematic procedure for finding the characteristics; this upper bound is always less than, or equal to, twice the lower bound of Hamming.[1]

Introduction

TWO notions are central to the class of error-correction codes to be considered here. One notion is that of error location by means of systematic parity checks,[1] and the other is that of generation of code letters by presuming they form an Abelian group.[2] These two notions, which have been proved equivalent,[2] may be used extensively in the study of error-correction codes which meet the following specifications:

1) The total number of binits in each code letter, including both information binits and redundant binits, is n.
2) The number of binits altered in a code letter during transmission is less than or equal to e.
3) If the number of redundant binits is r, then every instance of error satisfying specification 2) must be uniquely indentified by a corrector[3] consisting of r binits.
4) Each redundant binit of a code letter is a linear sum (modulo 2) of one or more information binits.

The Linear Independence of Characteristics

A parity-check code is completely specified by its characteristics[4] which are actually a tabulation of the decoding relations. If each code letter consists of n binits, r of which are redundant, then the code is described by n characteristics, each of which is an r-binit number. The essential nature of the characteristics associated with a given code lies in the degree to which they are linearly independent. If e is the maximum number of binits which may be altered during transmission of a code letter, then any set of not more than $2e$ characteristics is linearly independent, as will be shown.

If the characteristics are regarded as r-dimensional vectors, all of whose components are zero or one, then addition of characteristics is performed according to the usual rules of vector addition with the appropriate modification that addition of components is performed (modulo 2). The kth decoding relation which produces the kth binit of the corrector is obtained from the kth components of the n characteristics. If the kth component of the jth characteristic is one, then the jth binit of the received code letter is included in the kth decoding relation. Thus, those binits (of the received code letter) whose sum (modulo 2) constitutes the kth parity check are singled out by the nonzero kth components of the n characteristics.

The r-encoding relations are obtained in a similar fashion from the n characteristics. The decoding relations applied to a code letter prior to transmission yield a corrector all of whose components are zero. If r of the n characteristics are chosen to be the unit vectors of r-dimensional space, then these characteristics will correspond to the redundant binits of the code letter while the remaining characteristics will correspond to the information binits. The nonzero kth components of the characteristics corresponding to the information binits will then single out those information binits of the code letter whose sum (modulo 2) is the kth redundant binit of the code letter.

The theorem concerning the characteristics of a parity check code may be stated as follows. The n-place, binary parity check code generated by n characteristics will correct all single, double, $\cdots$, and e-tuple errors if and only if every set of $2e$ characteristics is linearly independent. For proof let $a_1, a_2, \cdots, a_n$ be the n binits of a typical code letter, and let $\vec{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jr})$ be the jth characteristic of the code. Then the following relations hold for the code letter prior to transmission.

$$\alpha_{11}a_1 \oplus \alpha_{21}a_2 \oplus \cdots \oplus \alpha_{n1}a_n = 0$$
$$\vdots$$
$$\alpha_{1r}a_1 \oplus \alpha_{2r}a_2 \oplus \cdots \oplus \alpha_{nr}a_n = 0.$$

The corrector $\vec{\gamma}$ is an r dimensional vector whose ith digit is obtained from

$$\gamma_i = \sum_{j=1}^{n} \alpha_{ji}a_j^1$$

where a_j^1 is the value of a_j after transmission. Suppose a_k is the only binit of the code letter which is altered during transmission; the corrector will then be the kth characteristic, $\vec{\alpha}_i$. Suppose $a_{i_1}, a_{i_2}, \cdots, a_{i_p}$ (where $p \leq e$) are the binits which are altered during transmission. Then the corrector is $\vec{\gamma} = \vec{\alpha}_{i_1} \oplus \vec{\alpha}_{i_2} \oplus \cdots \vec{\alpha}_{i_p}$. Two distinct, allowable instances of error will have the same corrector if and

* Manuscript received by the PGIT, May 2, 1958. The research reported in this paper was supported in part by the AF Cambridge Res. Center under Contract No. AF19(604) 1836.

† 520 E. Buffalo St., Ithaca, N. Y.

[1] R. Hamming, "Error detecting and error correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; April, 1950, (See p. 160).

[2] D. Slepian, "A class of binary signaling alphabets," *Bell Sys. Tech. J.*, vol. 35, pp. 203–234; January, 1956.

[3] M. Golay, "Binary coding," IRE Trans. on Information Theory, no. PGIT-4, pp. 23–28; September, 1954.

[4] The notion of characteristic employed here is slightly more general than that of Golay, *ibid.*

only if some set of not more than $2e$ characteristics is linearly dependent.

To accomplish decoding, a complete decoding book is necessary. This is simply a list of the allowable instances of error and the associated unique correctors; the list of correctors consists of all possible linear sums of characteristics taken not more than e at a time. The characteristics for the case $n = 32$ and $e = 2$ are given in the final section.

Upper and Lower Bounds on r

Hamming[1] obtained a lower bound on the minimum number of redundant binits required to correct up to and including e errors in a parity-check code where each code letter consists of n binits:

$$2^r \geq \sum_{i=0}^{e} \binom{n}{i}.$$

An upper bound on r may be obtained by considering a crude but systematic procedure for finding the characteristics of a code for fixed n and e.

The first characteristic is chosen arbitrarily, subject only to the condition that it not be the null vector. The second characteristic is chosen so that it is different from the first and the null vector. The third characteristic is chosen so that it is different from the first two, from the null vector, and from the sum of the first two characteristics. The kth characteristic is chosen so that it is different from the previously chosen characteristics, from the null vector, and from all m-tuple sums of previously chosen characteristics where m is any positive integer less than or equal to $2e - 1$. The whole computation may be tabulated as follows.

1) $\vec{\alpha}_1 \neq \vec{0}$

2) $\vec{\alpha}_2 \neq \vec{0}, \vec{\alpha}_1$

3) $\vec{\alpha}_3 \neq \vec{0}, \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_1 \oplus \vec{\alpha}_2$

4) $\vec{\alpha}_4 \neq \vec{0}, \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3, \vec{\alpha}_1 \oplus \vec{\alpha}_2, \vec{\alpha}_1 \oplus \vec{\alpha}_3, \vec{\alpha}_2 \oplus \vec{\alpha}_3$

$\vdots$

$n)$ $\vec{\alpha}_n \neq 0, \vec{\alpha}_j \qquad j = 1, 2, \cdots, n-1$

$\neq \vec{\alpha}_{j_1} \oplus \vec{\alpha}_{j_2}$

$\vdots$

$\vec{\alpha}_n \neq \vec{\alpha}_{j_1} \oplus \vec{\alpha}_{j_2} \oplus \cdots \oplus \vec{\alpha}_{j_{2e-1}}$.

Each stage of the computation contains all the entries of the previous stages. This method of generation guarantees the needed independence of the characteristics. The method will not go to completion successfully unless r (the dimensionality of the characteristics) is sufficiently large. To find such an r, consider the worst possible outcome of making arbitrary choices for the α's. That outcome consists of all the entries in the nth stage being distinct; to insure success in the face of this worst outcome, 2^r need only be greater than the total number of entries in the nth stage.

Table I
Characteristic Vectors for Double-Error Correction of 32-Binit Messages

i	αi			
1	1 0 0 0	0 0 0 0	0 0 0 0	0
2	0 1 0 0	0 0 0 0	0 0 0 0	0
3	0 0 1 0	0 0 0 0	0 0 0 0	0
4	0 0 0 1	0 0 0 0	0 0 0 0	0
5	0 0 0 0	1 0 0 0	0 0 0 0	0
6	0 0 0 0	0 1 0 0	0 0 0 0	0
7	0 0 0 0	0 0 1 0	0 0 0 0	0
8	0 0 0 0	0 0 0 1	0 0 0 0	0
9	0 0 0 0	0 0 0 0	1 0 0 0	0
10	0 0 0 0	0 0 0 0	0 1 0 0	0
11	0 0 0 0	0 0 0 0	0 0 1 0	0
12	0 0 0 0	0 0 0 0	0 0 0 1	0
13	0 0 0 0	0 0 0 0	0 0 0 0	1
14	1 1 0 0	1 1 0 0	0 0 0 0	0
15	1 1 0 0	0 0 1 1	0 0 0 0	0
16	1 1 0 0	0 0 0 0	1 1 0 0	0
17	1 1 0 0	0 0 0 0	0 0 1 1	0
18	0 1 1 0	0 1 1 0	0 0 0 0	0
19	0 1 1 0	0 0 0 1	1 0 0 0	0
20	0 1 1 0	0 0 0 0	0 1 1 0	0
21	0 1 1 0	0 0 0 0	0 0 0 1	1
22	1 0 0 1	1 1 1 0	0 0 0 0	0
23	1 0 0 1	0 0 1 1	1 0 0 0	0
24	1 0 0 1	0 0 0 0	1 1 1 0	0
25	1 0 0 1	0 0 0 0	0 0 1 1	1
26	1 0 0 1	1 0 0 1	0 0 1 0	0
27	1 0 0 1	0 1 0 0	1 0 0 1	0
28	1 0 0 1	0 0 1 0	0 1 0 0	1
29	1 0 0 1	1 0 0 0	1 0 0 0	1
30	1 0 0 1	0 1 0 1	0 1 0 0	0
31	1 0 0 1	1 1 0 1	0 0 0 1	1
32	1 1 1 1	1 1 1 1	1 1 1 1	1

$$2^r > \sum_{i=0}^{2e-1} \binom{n-1}{i}.$$

The smallest integer r that satisfies this last equation is an upper bound on the minimum number of redundant binits required to correct not more than e errors in a code letter consisting of n binits. If $\{a\}$ denotes the smallest integer greater than or equal to a, then[5]

$$r_{\max} = \left\{\log_2 \left[1 + \sum_{i=0}^{2e-1} \binom{n-1}{i}\right]\right\}$$

$$r_{\min} = \left\{\log_2 \left[\sum_{i=0}^{e} \binom{n}{i}\right]\right\}.$$

As an example consider a 32-binit code letter in which all single, double, and triple errors are to be corrected.

$$r_{\max} = \left\{\log_2 \left[1 + \sum_{i=0}^{5} \binom{31}{i}\right]\right\} = 18$$

$$r_{\min} = \left\{\log_2 \left[\sum_{i=0}^{3} \binom{32}{i}\right]\right\} = 13.$$

It has been found[6] that the required number of redundant binits is at least 15.

[5] A similar but less fine upper bound is given by E. Gilbert, "A comparison of signaling alphabets," *Bell Sys. Tech. J.*, vol. 31 pp. 504–522; May, 1952.

[6] I. Reed, "A class-of multiple-error-correcting codes and the decoding scheme," IRE Trans. on Information Theory, no. PGIT-4, pp. 38–49; September, 1954.

The upper bound is useful only when e is small compared to n. In that case $r_{max} - r_{min}$ will be small compared to n. This is shown in the following argument in which L means $\lim_{n\to\infty}$.

$$L\left[\log_2 \frac{\sum_{i=0}^{2e-1}\binom{n-1}{i}}{\sum_{i=0}^{e}\binom{n}{i}} + \log_2 \frac{\binom{n}{e}}{\binom{n-1}{2e-1}}\right] = 0$$

$$L\left[\log_2 \frac{\binom{n}{e}n^{e-1}e!}{\binom{n-1}{2e-1}(2e-1)!}\right] = 0$$

$$L\,\frac{r_{max} - r_{min}}{n} = L\,\frac{(e-1)\log_2 n - \log_2\left[\frac{(2e-1)!}{e!}\right]}{n}$$

$$L\,\frac{r_{max} - r_{min}}{n} = 0.$$

A more interesting property of the upper bound is that it is always less than or equal to twice the lower bound.

$$\binom{n}{i+j} \le \binom{n}{i}\binom{n}{j}$$

$$\sum_{i=0}^{2e}\binom{n}{i} \le \sum_{i=0}^{e}\binom{n}{i}^2 + \sum_{\substack{i,j=0\\ i\neq j}}^{e}\binom{n}{i}\binom{n}{j}$$

$$\sum_{i=0}^{2e}\binom{n}{i} \le \left[\sum_{i=0}^{e}\binom{n}{i}\right]^2$$

$$\frac{r_{max}}{r_{min}} \le 2.$$

Miscellaneous Remarks

Experience has shown when e is small and n not too large, the characteristics may be readily found by trial and error with the help of the bounds on r. The case for $n = 32$ and $e = 2$ was worked by hand and is given in Table I on the preceding page. It was found that $r_{min} = 10$ and $r_{max} = 13$. The choice of $\alpha_1 - \alpha_{13}$ was immediate, and the rest followed with some labor. The complete decoding book would consist of Table I and 496 other entries.

The Utility of a Communication Channel and Applications to Suboptimal Information Handling Procedures*

MICHAEL B. MARCUS†

Summary—**This paper demonstrates the applicability of the functional equations of dynamic programming to information theory problems. Yielding the same results as those obtainable by Shannon's equations, the functional equations can be modified also to consider the many restrictions enforced upon information systems by the real world. A result of the application of functional equations to systems operating under suboptimum conditions is that the information rate of a system is dependent upon the manner in which the information is used.**

The Kelly concept—the gain of a gambler who wagers his capital on the outcome of a communication channel—is used to determine the information rate of the channel. The mathematical analysis follows the stochastic multistage decision process technique of Bellman and Kalaba. Together with some extensions by the author, the Kelly-Bellman-Kalaba model of communication is repeated. The models are analyzed for the optimum case and examined for various suboptimum conditions. The gambler's betting policy is analogous to information usage; restrictions upon this policy affect the information rate of the system. They can require that the policy which is best under optimum conditions be replaced by other policies which, although inferior in the ideal case, are better able to compensate for the restrictions.

A null zone reception system first analyzed by Bloom and others is reanalyzed to provide a concrete example of the latitude of operation allowed by the functional equation approach. Bloom's analysis assumed that the system operates under optimum conditions. His results are duplicated, and their expression indicates their alteration by suboptimum conditions. An appendix expresses the results of this paper in the form used by Bloom.

* Manuscript received by the PGIT, April 29, 1958.
† The Rand Corp., Santa Monica, Calif.

Introduction

THE problem of determining the utility of a communication channel as one within the framework of the theory of multistage decision processes, or dynamic programming, has been discussed by Bellman and Kalaba [1].

To introduce the utility concept, we consider a gambler who uses the output of a communication channel in connection with the determination of a betting policy that will maximize the expected value of the logarithm of his capital after N stages of betting. The amount of capital that the gambler can acquire using communication systems with various properties gives us a quantitative measure by which these systems can be compared.

In practice, we are often given a fixed communication channel and the task of deciding how it can be used most efficiently. These problems can be treated by maximizing Shannon's equations [3] for the rate of a channel in bits per symbol.

We see that the utility concept employed in a dynamic programming approach yields the same results that are obtained by using Shannon's equation. In addition, the equations of dynamic programming are easy to formulate and admit a simple solution via digital computers. Often, in fact, elementary analytic solutions may be obtained. These equations are also easy to interpret—a characteristic usually absent in information theory due to a number of complexities inherent in its methods.

We also consider the effects of placing certain restrictions upon the manner in which the information taken from the communication system can be utilized. By doing this, it is seen that the information content of a system is dependent upon the way in which the information is used. For communications components with variable parameters, we see that the setting of a component for a theoretically optimal utilization of information is not necessarily the best setting if we are forced to operate in some suboptimal situations. Dynamic programming techniques are used in these demonstrations. These techniques and the utility concept add a breadth to information theory that enables it to be adapted to many realistic situations.

Two problems are considered in this paper. The first one is a hypothetical gambling situation that is used to demonstrate the above assertions. The second problem is a reanalysis and extension of a communication problem solved by Bloom, *et al.* [2]. This is used to show the applicability of these ideas to a concrete situation. In both cases, we show that some theoretically optimal procedures are not as good as some suboptimal procedures when certain restrictions are placed upon the manner in which the information obtained from the system can be employed.

The work on the reanalysis of Bloom's paper [2] was done with Bellman, Kalaba, and M. Juncosa. Their ideas and opinions were also very helpful in the formation of the remainder of the paper.

The Gambling Problem

In the first problem, the reasoning follows that given in Bellman and Kalaba [1], [5].

We are presented with the following situation. The *a priori* probability that a future event will be successful is 1/2. A gambler has a communication system on which he receives a positive pulse if the event is successful and a negative pulse if the event is not successful. However, because the channel is noisy, the gambler only knows with probability p that if a positive pulse is received, then a positive pulse has been transmitted. On the basis of this information, the gambler wagers a fraction of his capital on the success or failure of each event and tries to maximize the expected value of the logarithm of his final capital after N stages of betting.

Consider now three ways in which the communication system might be employed.

1) The gambler could require one pulse for each event, and bet on the information conveyed by this pulse.
2) The gambler could require two pulses for each event, and bet only if both pulses (each of which has a probability p of being correct) agree. In this case, between two and four pulses are the expected number required for each bet.
3) The gambler could require two pulses for each event and add them (symmetrical, additive, Gaussian noise will be assumed). The policy employed here is similar to the policy employed in Case 1) except that the gambler would have a greater assurance of the outcome of the event, but would pay for it by requiring two pulses for each bet.

Mathematical Formulation

The methods of dynamic programming will be used. We shall determine the gambler's gain if optimal policies are followed in each of the above processes. The analysis pertaining to Case 1) can be found in Bellman and Kalaba [1]; it will be repeated here for the sake of completeness of this report.

We define for all three cases,

$f_N(x)$ = the expected value of the logarithm of the gambler's capital after N pulses have been transmitted when the initial capital is x.

Let y $(0 \leq y \leq x)$ be the amount that the gambler wagers at each stage of betting. Then in Case 1), we have

$$f_N(x) = \max_{0 \leq y \leq x} [pf_{N-1}(x + y) + qf_{N-1}(x - y)]. \quad (1)$$

$$f_1(x) = \max_{0 \leq y \leq x} [p \log (x + y) + q \log (x - y)], \quad (2)$$

where $p + q = 1$.

The value of y that maximizes f_1 is $y = (p - q)x$ for $p \geq q$ and $y = 0$ for $p < q$. In these problems, we assume $p > q$. If p were less than q we would bet that the event was not successful; the case for $p = q$ is trivial since we know this much before the pulse is sent. Substituting for y in (2), we get

$$f_1 = \log x + [\log 2 + p \log p + q \log q] \quad (3)$$

and by induction we can show that

$$f_N(x) = \log x + N(\log 2 + p \log p + q \log q). \quad (4)$$

In Case 2), the gambler, when he bets, will win with probability $p^2/(p^2 + q^2)$ and lose with probability $q^2/(p^2 + q^2)$. However, recall that two pulses are required for each decision and with probability $2pq$ the two pulses will not agree and the gambler will not be able to make a bet. Taking these facts into account, the recurrence relations are

$$f_N(x) = \operatorname{Max}_{0 \le y \le x} \left[\frac{p^2}{p^2 + q^2} f_{N-2}(x + y) + \frac{q^2}{p^2 + q^2} f_{N-2}(x - y) + 2pq f_{N-2}(x) \right], \quad (5)$$

$$f_2(x) = \operatorname{Max}_{0 \le y \le x} \left[\frac{p^2}{p^2 + q^2} \log (x + y) + \frac{q^2}{p^2 + q^2} \log (x - y) + 2pq \log x \right]. \quad (6)$$

The value of y that maximizes f_2 is

$$y = \frac{(p^2 - q^2)x}{p^2 + q^2}$$

for $p \geq q$, and $y = 0$ for $p < q$. Recall, we assume $p > q$. Substituting y in (6) and proceeding inductively, we get

$$f_N(x) = \log x + \frac{N}{2} [p^2 \log 2p^2 + q^2 \log 2q^2 - (p^2 + q^2) \log (p^2 + q^2)]. \quad (7)$$

In Case 3), the functional equations are identical to the equations of Case 1), except that now N is replaced by $N/2$ and the value for p is different. The new value for p will be denoted by p_1. p_1 is the value, for the probability that the polarity of the pulse received is the same as the polarity of the pulse that was transmitted, that is obtained if we assume a Gaussian distribution for p and decrease its variance by $1/\sqrt{2}$. Clearly, $p_1 > p$ since in this case we are using a sample of two pulses to determine the outcome of an event whereas in Case 1) we only used a sample of one pulse. Therefore in Case 3),

$$f_N(x) = \log x + \frac{N}{2} (\log 2 + p_1 \log p_1 + q_1 \log q_1). \quad (8)$$

These equations were derived by betting the amount y that optimized $f_N(x)$. The values of $f_N(x)$ yield a quantitative measure for the utility of the communication channel for the three systems of betting. However, suppose that practical circumstances prohibit the assumptions made in the formulation of the three equations to be satisfied. These unavoidable restrictions can significantly alter the relative merits of the systems. This problem is considered in the next section.

Discussion

From (1) the expected value of the increase in the logarithm of the capital per pulse is $(\log 2 + p \log p + q \log q)$; this is equivalent to the maximum rate of transmission in bits per symbol and is the same value that would be obtained using Shannon's formula [1], [7]. Since this is the expression for the maximum rate in bits per symbol, it cannot be more than doubled by sending two symbols. Therefore $f_N(x)$ for Case 1) is equal to or greater than $f_N(x)$ for Case 2) or Case 3). Actually, they are equal only when $p = q$ and this is trivial since the gambler bets only when $p > q$. By computing the values for $f_N(x)$ for the various processes, we find that $f_N(x)$ for Case 1) is greater for all $p > q$ than $f_N(x)$ for Case 3) and that $f_N(x)$ for Case 3) is greater for all $p > q$ than $f_N(x)$ for Case 2).

The processes are in the order of 1), 3), and 2), with respect to the magnitude of the rate of transmission in bits per symbol. However, for the above results to hold, it must be true that two pulses are twice as expensive as one pulse. If they are not, then Case 2) and Case 3) are favored. Also, it has been assumed that the gambler can bet as often as he pleases and that there is no expense in making a bet. Neither of these assumptions is necessarily correct. If the gambler wins too often, he will find it difficult to place bets; if he is wagering at a casino, he must pay for each bet that he places. These factors strongly favor Case 2) and they also favor Case 3). Of course, as $p \to 1$ Case 1) becomes favorable since if there is a high probability of winning, the gambler would want to bet as often as possible.

At all times, we must consider the gambler and his goal, which is to maximize the expected value of the logarithm of his capital. If there are any restrictions placed upon the gambler's behavior, then they must be included in the functional relationships. Only in this way can we determine which of the policies will best serve the gambler in the given situation.

It is difficult to perceive the realism of the gambling problem, but the same analysis and comments apply to any system in which pulses are received, a decision is made on the basis of these pulses, and a quantity of available resources utilized according to this decision.

These factors have obvious counterparts in economic processes.

The object of this discussion is to show that the utility of the communication channel is influenced by factors external to it, and that, *only by considering the manner in which the information will be utilized*, can the communication system be evaluated.

Threshold Settings of a Null Zone in Binary Pulse Transmission

The problem studied by Bloom [2] will now be reanalyzed using the foregoing ideas. Bloom shows that for binary transmission with symmetrical, additive, Gaussian noise, the rate of transmission in bits per symbol is increased if a threshold above the zero level is set in the receiver and all signals that are received below this threshold are rejected. A curve is drawn for the setting of the threshold as a function of the signal-to-noise ratio of the transmitted pulses. Since this curve is based on Shannon's equation for channel capacity [3], it is assumed that the optimal code is used, or, translated into the gambling problem, that the gambler is operating under ideal conditions. We will show that if restrictions are placed upon the gambler, it can be more profitable for him to use a threshold setting other than the one that he would use under ideal circumstances.

It is assumed that the transmitted pulses are positive and negative with equal amplitudes and that the noise is symmetrical, additive Gaussian. Again we consider the gambler, but this time he can only determine whether a pulse is received above the positive threshold or below the negative threshold, that he has set in his receiver. Consequently, the threshold setting alone determines the probability p that the polarity of the pulse that is received is the same as the polarity of the pulse that was transmitted. (Recall that in Case 1), the gambler could determine the amplitude of the received pulse and thereby estimate a precise probability for each pulse that he receives.)

Consider that the amplitude of the pulses are $\pm \sqrt{P}$. Then we have the condition of Fig. 1.

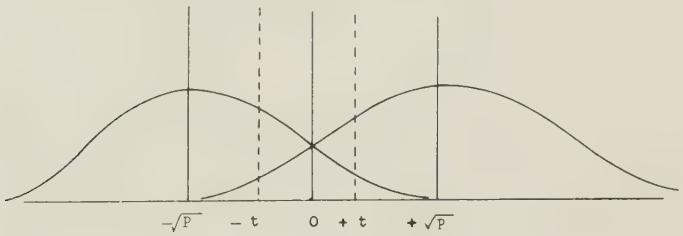


Fig. 1.

The gambler sets the thresholds at $\pm t$ and bets according to the following policy:

1) If the signal $< -t$, bet y that a negative pulse has been transmitted.
2) If $-t <$ signal $< +t$, do not bet.
3) If the signal $> t$, bet y that a positive pulse has been transmitted.

By virtue of symmetry, we need only consider the positive half of the graph. Let ${}_{a}\overset{+}{A}_{b}$ be the area from a to b under the Gaussian curve with positive mean.

Now, as in the previous example, we define

$f_N(x)$ = the expected value of the logarithm of the gambler's capital after N pulses have been transmitted.

If a positive pulse has been transmitted, the gambler will receive a pulse above the positive threshold with a probability $p = {}_{t}\overset{+}{A}_{\infty}$. Observe that even if a negative pulse has been transmitted, with probability $q = {}_{t}\overset{-}{A}_{\infty}$, the gambler will receive a signal above the positive threshold. (It is clear that in the former case he would bet and win, whereas in the latter case he would bet and lose.) Furthermore, with probability $({}_{0}\overset{+}{A}_{t} + {}_{0}\overset{-}{A}_{t})$ the amplitude of the pulse will lie below the threshold level and no signal will be received. Recall that x is the gambler's capital at each stage of betting and y is the amount that he bets.

From these conditions we can write the functional equations,

$$f_N(x) = \operatorname*{Max}_{\substack{0 \le y \le x \\ 0 \le t}} [{}_{t}\overset{+}{A}_{\infty} f_{N-1}(x+y) + {}_{t}\overset{-}{A}_{\infty} f_{N-1}(x-y) + ({}_{0}\overset{+}{A}_{t} + {}_{0}\overset{-}{A}_{t}) f_{N-1}(x)], \tag{9}$$

and for the first stage,

$$f_1(x) = \operatorname*{Max}_{\substack{0 \le y \le x \\ 0 < t}} [{}_{t}\overset{+}{A}_{\infty} \log (x+y) + {}_{t}\overset{-}{A}_{\infty} \log (x-y) + ({}_{0}\overset{+}{A}_{t} + {}_{0}\overset{-}{A}_{t}) \log x]. \tag{10}$$

This can be rewritten as

$$f_1(x) = \operatorname*{Max}_{\substack{0 \le y \le x \\ 0 \le t}} \left[{}_{t}\overset{+}{A}_{\infty} \log \frac{x+y}{x} - {}_{t}\overset{-}{A}_{\infty} \log \frac{x}{x-y} + \log x \right]. \tag{11}$$

From this equation we can see that it is always advantageous to have a threshold set in the receiver (*i.e.*, $t > 0$). For the zero threshold case, $t = 0$. We will take $f_1(x)$ at $t = 0$ and subtract it from a value for $f_1(x)$ at $t = t'$. As t' approaches zero, this difference is positive, thus a threshold $t > 0$ increases the magnitude of $f_1(x)$.

$$\underset{t=t'}{f_1(x)} - \underset{t=0}{f_1(x)} = {}_{0}\overset{-}{A}_{t} \log \frac{x}{x-y} - {}_{0}\overset{+}{A}_{t} \log \frac{x+y}{x}. \tag{12}$$

For t' near zero, ${}_{0}\overset{-}{A}_{t} = {}_{0}\overset{+}{A}_{t} = K$, and

$$\underset{t=t'}{f_1(x)} - \underset{t=0}{f_1(x)} = K\left[\log \frac{x}{x-y} - \log \frac{x+y}{x} \right], \tag{13}$$

and since $x/(x-y) > (x+y)/x$ $(0 < y < x)$, this value is positive.

From (11) we can show by induction that

$$f_N(x) = \log x + \operatorname*{Max}_{\substack{0 \le y \le x \\ 0 \le t}} N\left[{}_{t}\overset{+}{A}_{\infty} \log \frac{x+y}{x}, - {}_{t}\overset{-}{A}_{\infty} \log \frac{x}{x-y} \right]. \tag{14}$$

If $f_N(x)$ is maximized with respect to y, *i.e.*,

$$y = \frac{({}_{t}\overset{+}{A}_{\infty} - {}_{t}\overset{-}{A}_{\infty})x}{({}_{t}\overset{+}{A}_{\infty} + {}_{t}\overset{-}{A}_{\infty})},$$

we have a value for the expected gain of the gambler, in N stages, as a function of t. Dividing this by N we get the gain per stage as a function of t and this is precisely the value that Bloom obtained for the rate in bits per symbol as a function of the threshold t (see Appendix). Thus, if we were to draw a graph for the value of t as a function of the signal-to-noise ratio, it would be identical to the graph compiled by Bloom. However, we must note that we used the value of y that maximizes $f_N(x)$. What if this value of y is not obtainable? In the case of the gambler this can occur if he were only allowed to wager certain

fixed amounts. Since $f_N(x)$ is both a function of t and y changing y requires that a new value of t be determined to maximize $f_N(x)$.

Regardless of the value of y, once it is fixed, a value of t can be determined that will make the gambler's gain as great as possible. Thus the curve in Bloom's paper should be replaced by a family of curves; the graph (for the case of the gambler) would be one for the threshold setting as a function of the signal-to-noise ratio and the amounts that can be wagered at each bet.

The significant factor is that when this system is set to operate most efficiently by assuming an optimal code, it operates most efficiently only if this optimal code is used. It is not uniformly efficient for *all* codes; in fact, the optimal setting of the threshold is dependent upon the type of code used.

Conclusion

When an information channel is included in a system, it becomes part of the system; its most efficient mode of operation is determined by the remainder of the system and this includes the manner in which the information is used.

A detailed analysis of these problems is possible through the use of the technique of dynamic programming. Although in some situations, the results obtained correspond to the results of Shannon's equations, the method can be extended to treat a number of contingent circumstances of the real world.

Appendix

We show that for an optimum betting policy, the results of the analysis of the null-zone reception problem are the same as those obtained by Bloom.

Eq. (14) for the expected value of the log of the gamblers capital after N pulses have been transmitted is

$$f_N(x) = \log x + \max_{\substack{0 \le y \le x \\ 0 \le t}} N\left[{}_t\overset{+}{A}_\infty \log \frac{x+y}{x} - {}_t\overset{-}{A}_\infty \log \frac{x}{x-y}\right]. \tag{15}$$

Differentiating (15) with respect to y and setting the result equal to zero, we see that the value of y that maximizes $f_N(x)$ is

$$y = \frac{{}_t\overset{+}{A}_\infty - {}_t\overset{-}{A}_\infty}{{}_t\overset{+}{A}_\infty + {}_t\overset{-}{A}_\infty}. \tag{16}$$

Substituting this value of y in (15) yields:

$$f_N(x) = \log x + N\left[{}_t\overset{+}{A}_\infty \log \frac{2\,{}_t\overset{+}{A}_\infty}{{}_t\overset{+}{A}_\infty + {}_t\overset{-}{A}_\infty} + {}_t\overset{-}{A}_\infty \log \frac{2\,{}_t\overset{-}{A}_\infty}{{}_t\overset{+}{A}_\infty + {}_t\overset{-}{A}_\infty}\right]. \tag{17}$$

It follows directly that the increase in the log of the gambler's capital per stage (or symbol) is

$${}_t\overset{+}{A}_\infty \log \frac{2\,{}_t\overset{+}{A}_\infty}{{}_t\overset{+}{A}_\infty + {}_t\overset{-}{A}_\infty} + {}_t\overset{-}{A}_\infty \log \frac{2\,{}_t\overset{-}{A}_\infty}{{}_t\overset{+}{A}_\infty + {}_t\overset{-}{A}_\infty}. \tag{18}$$

If, instead of "log of capital" we consider "information," then (18) is the rate of information in bits per symbol.

Bloom places all signals falling above the positive threshold in class x_1. He then defines $p\ (I \mid x_1)$ and $p\ (\mathrm{III} \mid x_1)$ as the conditional probabilities that given a signal in class x_1 it is positive and negative, respectively. Symmetry allows him to consider only the positive half of the range. It is apparent that $p(I \mid x_1) = {}_t\overset{+}{A}_\infty$ and the $p\ (\mathrm{III} \mid x_1) = {}_t\overset{-}{A}_\infty$. Employing Shannon's equation for the information rate of a system in bits per symbol, Bloom obtains

$$p(\mathrm{I} \mid x_1) \log \frac{2p(\mathrm{I} \mid x_1)}{p(\mathrm{I} \mid x_1) + p(\mathrm{III} \mid x_1)} + p(\mathrm{III} \mid x_1) \log \frac{2p(\mathrm{III} \mid x_1)}{p(\mathrm{I} \mid x_1) + p(\mathrm{III} \mid x_1)}. \tag{19}$$

This is identical to (18).

Bibliography

[1] R. Bellman and R. Kalaba, "On the role of dynamic programming in statistical communication theory," IRE Trans. on Information Theory, vol. IT-3, pp. 197–203; September, 1957.

[2] F. J. Bloom, S. S. L. Chang, B. Harris, A. Hauptschein, and K. C. Morgan, "Improvement of binary transmission by null zone reception," Proc. IRE, vol. 45, pp. 963–975; July, 1957.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell. Sys. Tech. J.*, vol. 27, pp. 379–433; July, 1948, and pp. 623–656; October, 1948.

[4] R. Bellman, "Dynamic Programming," Princeton University Press, Princeton, N. J.; 1957.

[5] R. Bellman and R. Kalaba, "Dynamic programming and statistical communication theory," *Proc. Natl. Acad. Sci. US*, vol. 43, pp. 749–751; August, 1957.

[6] J. Kelly, "A new interpretation of information rate," *Bell Sys. Tech. J.*, vol. 35, pp. 917–926; July, 1956.

[7] A. Wald, "Sequential Analysis," John Wiley and Sons, Inc., New York, N. Y., 1947.

[8] J. J. Bussgang and D. Middleton, "Optimum sequential detection of signals in noise," IRE Trans. on Information Theory, vol. IT-1, pp. 5–18; December, 1955.

[9] H. Blasbalg, "Bernoulli Sequential Filtering," Radiation Lab., The Johns Hopkins University, Baltimore, Md., Tech. Rep. No. AF-27; January, 1956.

Capacity of a Certain Asymmetrical Binary Channel with Finite Memory*

SZE-HOU CHANG†

Summary—The capacity of a certain asymmetrical binary channel is studied under the following conditions. 1) Blocks of equal numbers of binary digits are used as the transmitting symbols. 2) The channel resumes its quiescent state at the beginning of each block. 3) The memory of the channel is characterized by the dependence of the noise probabilities for each digit upon the preceding digit or digits in the same block. It is shown that, by means of simple rules and with the aid of a single set of curves or a table, the calculation of the capacity can be reduced to a routine process.

Introduction

In recent years the binary channel has become increasingly important in the storage and transmission of information. There are numerous developments of physical devices for use in such channels, as well as theoretical investigations[1–4] of binary coding and detection schemes. One of the most important criteria upon which the design of these devices and schemes is based is the channel capacity as defined by Shannon.[5]

A symmetrical binary channel, abbreviated as SBC, is a channel in which the noise probabilities[6] are identical for the two digits, 0 and 1, as indicated by Fig. 1. The capacity of an SBC is well known,[5] and is given by

$$C = 1 + \alpha \log \alpha + (1 - \alpha) \log (1 - \alpha) \text{ bit}, \qquad (1)$$

where the log is taken to the base 2.

While the maximization procedure used in deriving (1) is relatively simple, it becomes rather involved when the asymmetrical binary channel (ABC) (see Fig. 2) or more general channels are considered. Muroga[7] has simplified the computation of the capacity in a certain class of discrete channels by introducing an auxiliary column matrix containing elements Q_i which are essentially equivalent to the logarithms of the noise probabilities. Using this technique, Silverman[8] has made a detailed analysis of the asymmetrical binary channel. However, his analysis deals with the case where the noise probabilities remain constant for successive digits, and therefore is applicable only to the asymmetrical binary channel with no memory, ABC (wnm).

The purpose of this paper is to use Muroga's method to compute the capacity of a certain asymmetrical binary channel with finite memory, ABC (wfm). The operation of this channel is as follows. First, the binary digits are transmitted in blocks having an equal number of digits. Second, there is sufficient time spacing to prevent interference between blocks so that the channel always resumes its quiescent state at the beginning of each block. Third, the memory of the channel is characterized by the dependence of the noise probabilities for each digit upon the preceding digit or digits in the same block.

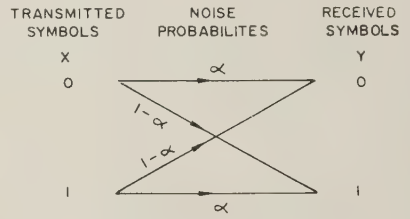


Fig. 1—Symmetrical binary channel.

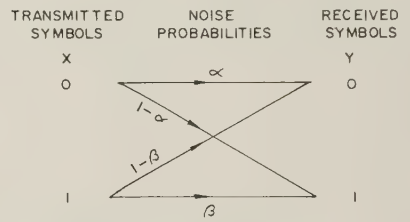


Fig. 2—Asymmetrical binary channel.

A block with n digits will be referred to as an n plet. The following discussion shows that the computation of the capacities of an ABC (wfm) using doublets, triplets, etc., is based upon the capacity computation of a single digit transmission, or more precisely, upon that of an ABC (wnm). Therefore, the latter case is reviewed first. In so doing, a set of curves and its associated table shall be introduced. These are not discussed in Silverman's paper but are found extremely useful in the following sections.

Asymmetrical Binary Channel (with No Memory)

The noise probability matrix of an ABC (wnm) is indicated by Fig. 2 and

* Manuscript received by the PGIT, June 5, 1958. This research was made possible through the support of the Electronics Res. Div. of the AF Cambridge Res. Center, Air Res. and Dev. Command.

† Northeastern University, Boston, Mass.

[1] R. W. Hamming, "Error detecting and error correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; April, 1950.

[2] I. S. Reed, "A class of multiple-error-correcting codes and the decoding scheme," IRE Trans. on Information Theory, no. PGIT-4, pp. 38–49; September, 1954.

[3] P. Elias, "Coding for Two Noisy Channels" in "Information Thoery," C. Cherry, ed., Thornton Butterworth, Ltd., London, Eng., pp. 61–74; 1956.

[4] D. Slepian, "A class of signaling alphabets," *Bell Sys. Tech. J.*, vol. 35, pp. 203–234; January, 1956.

[5] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communications," University of Illinois Press, Urbana, Ill.; 1949.

[6] Noise probabilities are the set of transitional probabilities $p(y \mid x)$ when x is transmitted and y is received.

[7] S. Muroga, "On the capacity of a discrete channel I," *J. Phys. Soc., Japan*, vol. 8, pp. 484–494; July-August, 1953.

[8] R. A. Silverman, "On binary channels and their cascades," IRE Trans. on Information Theory, vol. IT-1, pp. 19–27; December, 1955.

$$A = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}. \tag{2}$$

The corresponding noise entropy matrix is given by

$$H] = \begin{bmatrix} H(\alpha) \\ H(\beta) \end{bmatrix} = \begin{bmatrix} -\alpha \log \alpha - (1-\alpha) \log (1-\alpha) \\ -(1-\beta) \log (1-\beta) - \beta \log \beta \end{bmatrix}. \tag{3}$$

Muroga introduced an auxiliary column matrix Q which is defined[9] as

$$Q] = -A^{-1}H]. \tag{4}$$

Upon the completion of matrix inversion and multiplication, the Q's are explicitly determined; thus,

$$Q] = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \frac{-1}{\alpha + \beta - 1} \begin{bmatrix} \beta H(\alpha) + (\alpha - 1)H(\beta) \\ (\beta - 1)H(\alpha) + \alpha H(\beta) \end{bmatrix}. \tag{5}$$

The capacity of an ABC (wnm) can be expressed in terms of the Q's by

$$C = \log \sum_{i=1}^{2} 2^{Q_i} = \log (2^{Q_1} + 2^{Q_2}). \tag{6}$$

This procedure was used by Silverman in the study of the ABC (wnm). Readers are referred to the original paper for the detailed graphs of C and other quantities as functions of α and β.[9] In the following discussion, similar procedures are used to obtain the capacity of the ABC (wfm). It then becomes evident that it is convenient to use Q_1 and Q_2 given by (5) as the building elements of more complex Q matrices which in turn determine the capacities of different channels. Therefore, it is desirable to calculate the values of Q_1 and Q_2 as functions of α and β. In Fig. 3, the contours of Q_1's are plotted in the α-β plane. Because of the inherent symmetry between the expressions of Q_1 and Q_2, the same contours may be used to read Q_2 by simply interchanging the coordinates α and β.

In practice, the values of α and β are more commonly found in the range of 0.5 to 1. Table I lists the computed values of Q_1 corresponding to this region of α and β with finer increments of both parameters.

To illustrate the use of the curves, the channel capacity of an ABC (wnm) as a function of α is computed for a few selected values of β by means of (6) and Fig. 3. Fig. 4[10] shows these results together with one additional curve, that of a symmetrical binary channel in which β is always set equal to α. It is interesting to note that the channel capacity is greater than zero except when $\alpha + \beta = 1$, in which case transmission of information is impossible because there is no change in the probabilities of the transmitted digits upon the reception of either of the two digits. In other words, the *a posteriori* probabilities

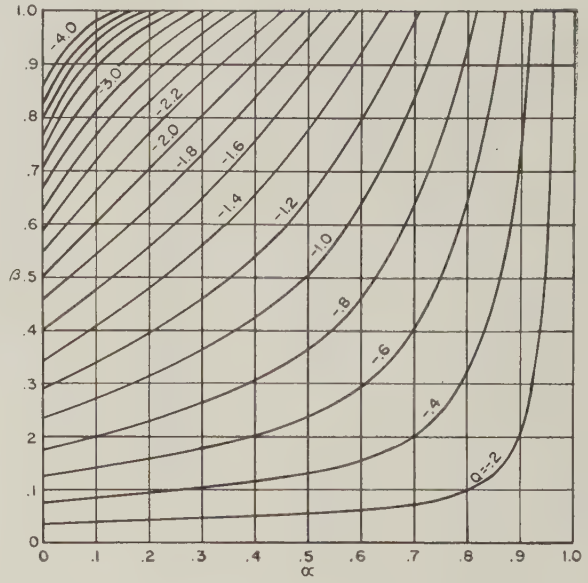


Fig. 3—Values of Q_1 for $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$. (Interchange α and β to read Q_2.)

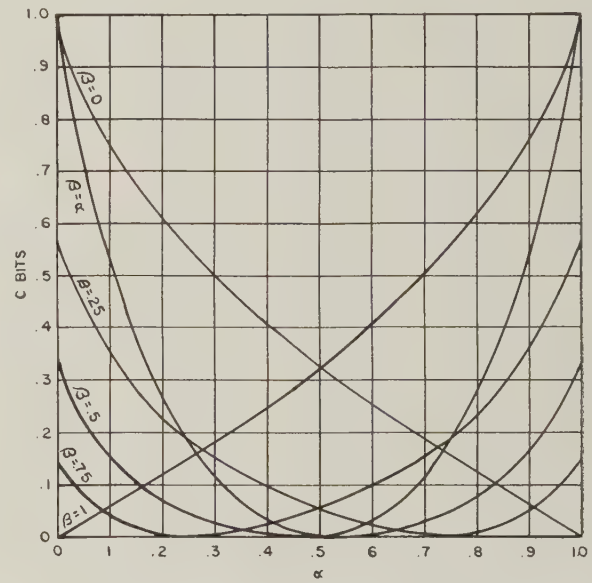


Fig. 4—Capacity of symmetrical and asymmetrical binary channel.

of the digits are the same as the *a priori* probabilities. Of further interest is the fact that in the vicinity of the zero-capacity point the capacity can be improved either by increasing α, the probability of correct recognition of 0, or by decreasing it. In either way the equivocation is reduced. It also should be noted that the full capacity of 1 bit per digit is reached when both α and β are equal to unity or when both are equal to zero. In the latter case an exchange of the roles of the two received digits results in a perfect channel.

Asymmetrical Binary Channel with Finite Memory for the Doublet

The simplest case of an ABC (wfm), one that used a doublet in a block, is studied now. Altogether there

[9] The notations in Muroga's and Silverman's papers are different from those used here.

[10] These curves were first obtained by R. D. Klein of Northeastern University using direct maximization independently of Silverman.

TABLE I*

VALUES OF Q FOR $0.5 \leq \left\{\begin{matrix}\alpha\\ \beta\end{matrix}\right\} \leq 1.0$

$\Delta\alpha = \Delta\beta = 0.02$

β \ α	0.50	0.52	0.54	0.56	0.58	0.60	0.62	0.64	0.66
0.50	−1.000000	−0.971125	−0.942225	−0.913225	−0.884087	−0.854750	−0.825175	−0.795296	−0.765056
0.52	1.028875	0.998845	0.968798	0.938667	0.908412	0.877966	0.847290	0.816318	0.784989
0.54	1.057775	1.026581	0.995377	0.964107	0.932720	0.901155	0.869369	0.837292	0.804866
0.56	1.086775	1.054392	1.022017	0.989586	0.957197	0.924357	0.891446	0.858256	0.824720
0.58	1.115912	1.082322	1.048753	1.015147	0.981453	0.947607	0.913559	0.879239	0.844583
0.60	1.145250	1.110425	1.075641	1.040838	1.005963	0.970950	0.935746	0.900282	0.864491
0.62	1.174825	1.138740	1.102719	1.066696	1.030619	0.994419	0.958041	0.921416	0.884474
0.64	1.204704	1.167331	1.130042	1.092776	1.057744	1.018061	0.980490	0.942682	0.904571
0.66	1.234944	1.196250	1.157666	1.119125	1.080566	1.041922	1.003131	0.964121	0.924818
0.68	1.265608	1.225558	1.185644	1.145798	1.105956	1.066048	1.026012	0.985772	0.945255
0.70	1.296772	1.255326	1.214044	1.172857	1.131698	1.090495	1.049183	1.007686	0.965926
0.72	1.328520	1.285633	1.242941	1.200372	1.157858	1.115323	1.072702	1.029915	0.986883
0.74	1.360946	1.316566	1.272416	1.228420	1.184508	1.140601	1.096632	1.052517	1.008179
0.76	1.394154	1.348225	1.302563	1.257089	1.211730	1.166405	1.121043	1.075561	1.029876
0.78	1.428273	1.380729	1.333494	1.286483	1.239622	1.192826	1.146023	1.099125	1.052048
0.80	1.463453	1.414220	1.365339	1.316725	1.268298	1.219972	1.171669	1.123300	1.074780
0.82	1.499880	1.448870	1.398262	1.347966	1.297900	1.247971	1.198102	1.148200	1.098176
0.84	1.537781	1.484893	1.432456	1.380392	1.328599	1.276987	1.225473	1.173963	1.122364
0.86	1.577446	1.522557	1.468188	1.414237	1.360613	1.307220	1.253969	1.200762	1.147504
0.88	1.619262	1.562226	1.506147	1.449812	1.394234	1.338940	1.283839	1.228828	1.173809
0.90	1.663756	1.604387	1.545687	1.487544	1.429856	1.372514	1.315422	1.258475	1.201567
0.92	1.711692	1.649785	1.588577	1.528044	1.468045	1.408466	1.349204	1.290150	1.231192
0.94	1.764267	1.699436	1.635480	1.572272	1.509692	1.447620	1.385947	1.324554	1.263329
0.96	1.823598	1.755397	1.688217	1.621913	1.556357	1.491420	1.426981	1.362917	1.299106
0.98	1.894333	1.821953	1.750784	1.680669	1.611464	1.543025	1.475223	1.407920	1.340987
1.00	2.000000	1.920855	1.843292	1.767119	1.692162	1.618250	1.545229	1.472942	1.401237

β \ α	0.68	0.70	0.72	0.74	0.76	0.78	0.80	0.82	0.84
0.50	−0.734391	−0.703227	−0.671479	−0.639054	−0.605846	−0.571726	−0.536546	−0.500120	−0.462219
0.52	0.753238	0.720990	0.688157	0.654646	0.620350	0.585136	0.548855	0.511317	0.472293
0.54	0.772021	0.738682	0.704760	0.670158	0.634769	0.598459	0.561075	0.522426	0.482280
0.56	0.790773	0.756334	0.721315	0.685617	0.649129	0.611718	0.573228	0.533466	0.492197
0.58	0.809521	0.773973	0.737848	0.701045	0.663453	0.624936	0.585335	0.544457	0.502063
0.60	0.828302	0.791632	0.754389	0.716472	0.677766	0.638134	0.597417	0.555417	0.511894
0.62	0.847142	0.809649	0.770964	0.731921	0.692091	0.651335	0.609492	0.566363	0.521706
0.64	0.866078	0.827121	0.787603	0.747420	0.706454	0.664563	0.621584	0.577318	0.531517
0.66	0.885440	0.845018	0.805285	0.762999	0.720881	0.677841	0.633715	0.588299	0.541346
0.68	0.904381	0.863062	0.821200	0.778686	0.735399	0.691195	0.645905	0.599327	0.551209
0.70	0.923825	0.881291	0.838224	0.794514	0.750039	0.704651	0.658182	0.610425	0.561129
0.72	0.943525	0.899748	0.855450	0.810521	0.764834	0.718242	0.670573	0.621618	0.571125
0.74	0.963531	0.918480	0.872923	0.826745	0.779821	0.731979	0.683106	0.632933	0.581222
0.76	0.983901	0.937541	0.890690	0.843233	0.795040	0.745959	0.695816	0.644398	0.591447
0.78	1.004703	0.956993	0.908810	0.860035	0.810539	0.760167	0.708742	0.656050	0.601829
0.80	1.026016	0.976908	0.927348	0.877213	0.826373	0.774671	0.721928	0.667926	0.612404
0.82	1.047935	0.997375	0.946385	0.894842	0.842610	0.789533	0.735428	0.680076	0.623213
0.84	1.070579	1.018503	0.966021	0.913010	0.859332	0.804826	0.749308	0.692559	0.634303
0.86	1.094094	1.040426	0.986381	0.931832	0.876640	0.820642	0.763651	0.705445	0.655753
0.88	1.118678	1.063323	1.007626	0.951455	0.894665	0.837102	0.778565	0.718832	0.657630
0.90	1.144593	1.087439	1.029979	0.972082	0.913601	0.854369	0.794194	0.732847	0.670052
0.92	1.172222	1.113119	1.053757	0.993999	0.933696	0.872688	0.810747	0.747948	0.683178
0.94	1.202154	1.140906	1.079453	1.017655	0.955358	0.892387	0.828545	0.763595	0.697255
0.96	1.236206	1.171745	1.107928	1.043829	0.979289	0.914129	0.848148	0.781104	0.712712
0.98	1.274291	1.207695	1.141055	1.074217	1.007018	0.939272	0.870771	0.801270	0.730478
1.00	1.329972	1.258987	1.188126	1.117224	1.046105	0.974573	0.902410	0.829362	0.755129

β \ α	0.86	0.88	0.90	0.92	0.94	0.96	0.98	1.00
0.50	−0.422554	−0.380738	−0.336243	−0.288308	−0.235732	−0.176404	−0.105666	−0.000000
0.52	0.431489	0.388515	0.342840	0.293694	0.239871	0.179287	0.107143	0.000000
0.54	0.440340	0.396213	0.349362	0.299014	0.243953	0.182045	0.108596	0.000000
0.56	0.449122	0.403845	0.355822	0.304277	0.247988	0.184807	0.110027	0.000000
0.58	0.457852	0.411423	0.362232	0.309495	0.251982	0.187539	0.111439	0.000000
0.60	0.466544	0.418963	0.368604	0.314675	0.255953	0.190245	0.112836	0.000000
0.62	0.475213	0.426477	0.374947	0.319828	0.259166	0.192929	0.114220	0.000000
0.64	0.483874	0.433978	0.381275	0.324964	0.263799	0.195599	0.115593	0.000000
0.66	0.492544	0.441481	0.387598	0.330090	0.267707	0.198258	0.116959	0.000000
0.68	0.501239	0.448999	0.393928	0.335218	0.271612	0.200911	0.118320	0.000000
0.70	0.509976	0.456547	0.400279	0.340358	0.275521	0.203564	0.119679	0.000000
0.72	0.518774	0.464143	0.406663	0.345520	0.279444	0.206223	0.121039	0.000000
0.74	0.527654	0.471802	0.413096	0.350716	0.283388	0.208894	0.122403	0.000000
0.76	0.536638	0.479546	0.419594	0.355960	0.287365	0.211583	0.123775	0.000000
0.78	0.545754	0.487396	0.426175	0.361266	0.291384	0.214298	0.125157	0.000000
0.80	0.555032	0.495378	0.432861	0.366651	0.295459	0.217048	0.126555	0.000000
0.82	0.564507	0.503524	0.439678	0.372135	0.299605	0.219841	0.127974	0.000000
0.84	0.574225	0.511869	0.446655	0.377744	0.303840	0.222691	0.129418	0.000000
0.86	0.584238	0.520461	0.453831	0.383506	0.308185	0.225611	0.130897	0.000000
0.88	0.594621	0.529361	0.461255	0.389460	0.312670	0.228622	0.132418	0.000000
0.90	0.605468	0.538648	0.468995	0.395660	0.317334	0.231747	0.133995	0.000000
0.92	0.616916	0.548438	0.477143	0.402179	0.322231	0.235024	0.135645	0.000000
0.94	0.629178	0.558909	0.485846	0.409131	0.327445	0.238507	0.137396	0.000000
0.96	0.642620	0.570370	0.495355	0.416714	0.333121	0.242291	0.139294	0.000000
0.98	0.658038	0.583489	0.506217	0.425355	0.339575	0.246583	0.141440	0.000000
1.00	0.679347	0.601546	0.521105	0.437151	0.348345	0.252387	0.144326	0.000000

*Interchange α and β to read Q_2. Values of Q's are negative or zero.

are four types of doublets, namely, 00, 01, 10, and 11. If the decision in the detection process is made on a block by block basis rather than on a digit by digit basis, then the channel is described by a 4 × 4 noise probability matrix associating the four transmitted doublets with the four received doublets, as shown below.

$$\begin{array}{c|c} x \backslash y & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \hline \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \end{array} \tag{7}$$

where $a_{ij} = p(y_j \mid x_i)$ is the conditional probability that the doublet y_j is received when the doublet x_i is transmitted. Because of the requirement that

$$\sum_{j=1}^{4} a_{ij} = 1, \tag{8}$$

only 12 out of the 16 parameters can be assigned independently.

Muroga's method as outlined in (4), (5), and (6) may be used to compute the channel capacity under this most general condition of doublet transmission. This by itself eliminates a considerable amount of computation as compared with the direct maximization process. However, the calculations are even more simplified if the decision at detection is made on a digit by digit basis and the channel memory is assumed to affect only the noise probabilities of the second digit.

Three noise probability matrices A_1, A_2', and A_2'' may be defined as follows.

For the first digit which is always sent during the quiescent state of the channel:

$$A_1 = \begin{bmatrix} \alpha_1 & 1-\alpha_1 \\ 1-\beta_1 & \beta_1 \end{bmatrix}. \tag{9}$$

For the second digit following "0":

$$A_2' = \begin{bmatrix} \alpha_2' & 1-\alpha_2' \\ 1-\beta_2' & \beta_2' \end{bmatrix}. \tag{10}$$

For the second digit following "1":

$$A_2'' = \begin{bmatrix} \alpha_2'' & 1-\alpha_2'' \\ 1-\beta_2'' & \beta_2'' \end{bmatrix}. \tag{11}$$

It is to be noted that these matrices are similar in structure to the one given by (2) for single digit blocks in the ABC (wnm). The matrix for the doublets as given above in (7) can be considered, in the present case, to be composed of the elements of these three submatrices in the following manner.

$$A = \begin{bmatrix} \alpha_1 \begin{bmatrix} \alpha_2' & 1-\alpha_2' \\ 1-\beta_2' & \beta_2' \end{bmatrix} & (1-\alpha_1) \begin{bmatrix} \alpha_2' & 1-\alpha_2' \\ 1-\beta_2' & \beta_2' \end{bmatrix} \\ (1-\beta_1) \begin{bmatrix} \alpha_2'' & 1-\alpha_2'' \\ 1-\beta_2'' & \beta_2'' \end{bmatrix} & \beta_1 \begin{bmatrix} \alpha_2'' & 1-\alpha_2'' \\ 1-\beta_2'' & \beta_2'' \end{bmatrix} \end{bmatrix}. \tag{12}$$

Thus, the probabilities that a transmitted doublet 00 is received as 00, 01, 10, and 11 are respectively $\alpha_1\alpha_2'$, $\alpha_1(1-\alpha_2')$, $(1-\alpha_1)\alpha_2'$, and $(1-\alpha_1)(1-\alpha_2')$. This matrix contains only six independent parameters, as compared with 12 in the general case given by (7).[11] It is to be expected that some simplification may be found by exploiting its special structure.

A useful way of decomposing the matrix of (12) is

$$A = \begin{bmatrix} A_2' & 0 \\ 0 & A_2'' \end{bmatrix} \begin{bmatrix} \alpha_1 & 1-\alpha_1 \\ 1-\beta_1 & \beta_1 \end{bmatrix} = \begin{bmatrix} A_2' & 0 \\ 0 & A_2'' \end{bmatrix} A_1. \tag{13}$$

Here the submatrices A_2' and A_2'' should be considered as if they were merely elements during the multiplication with the elements of A_1, and expanded thereafter if desired. Using this method, the inversion of A as required in (4) is much simplified:

$$A^{-1} = A_1^{-1} \begin{bmatrix} A_2'^{-1} & 0 \\ 0 & A_2''^{-1} \end{bmatrix}. \tag{14}$$

It is to be noted that the inversion of the 4 × 4 matrix is effected by the inversion of three 2 × 2 matrices all of which are similar in nature, for example,

$$A_1^{-1} = \frac{1}{\alpha_1 + \beta_1 - 1} \begin{bmatrix} \beta_1 & \alpha_1 - 1 \\ \beta_1 - 1 & \alpha_1 \end{bmatrix} = \begin{bmatrix} \gamma_1 & 1-\gamma_1 \\ 1-\partial_1 & \partial_1 \end{bmatrix}. \tag{15}$$

The last step also defines two new parameters γ_1 and ∂_1. The inverses of A_2' and A_2'' are obtained in the same manner and result in similar parameters, γ_2', ∂_2' and γ_2'', ∂_2''.

In order to form the auxiliary matrix Q, as required in (4), the noise entropy matrix should be computed first. This is easily done from (12) by computing on a row by row basis. After simplification, we get

$$H] = \begin{bmatrix} H(\alpha_1) + H(\alpha_2') \\ H(\alpha_1) + H(\beta_2') \\ H(\beta_1) + H(\alpha_2'') \\ H(\beta_1) + H(\beta_2'') \end{bmatrix} \tag{16}$$

where all elements are defined in exactly the same way as in (3) except for appropriate subscripts and superscripts. Finally, the matrix Q is computed.

[11] It can be shown that the ratio of the two numbers is 2^{n-1} for the n plet.

$$Q] = -A^{-1}H]$$

$$= \begin{bmatrix} \begin{bmatrix} Q_{11} \\ Q_{11} \end{bmatrix} + \gamma_1 & \begin{bmatrix} Q'_{21} \\ Q'_{22} \end{bmatrix} + (1-\gamma_1) & \begin{bmatrix} Q''_{21} \\ Q''_{22} \end{bmatrix} \\ \begin{bmatrix} Q_{12} \\ Q_{12} \end{bmatrix} + (1-\partial_1) & \begin{bmatrix} Q'_{21} \\ Q'_{22} \end{bmatrix} + \partial_1 & \begin{bmatrix} Q''_{21} \\ Q''_{22} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}. \quad (17)$$

In this expression the submatrices $\begin{bmatrix} Q_{11} \\ Q_{12} \end{bmatrix}$, $\begin{bmatrix} Q'_{21} \\ Q'_{22} \end{bmatrix}$, and $\begin{bmatrix} Q''_{21} \\ Q''_{22} \end{bmatrix}$ are functions of A_1, A'_2, and A''_2, respectively, in the same way that the matrix $\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}$ is the function of A in (5) of ABC (wnm). Therefore, the values of their elements can be read directly from Fig. 3 or Table I with appropriate values of α and β. The submatrices $\begin{bmatrix} Q_{11} \\ Q_{11} \end{bmatrix}$ and $\begin{bmatrix} Q_{12} \\ Q_{12} \end{bmatrix}$ are formed by iterating the elements of $\begin{bmatrix} Q_{11} \\ Q_{12} \end{bmatrix}$. An explanation of the notation is perhaps in order. The first subscript and the superscript, if any, of each Q_{ij} denotes the correspondence to the original matrices, *i.e.*, the A's, which generate them. The second subscript, 1 or 2, denotes the position of the Q_{ij} in the submatrice $\begin{bmatrix} Q_{i1} \\ Q_{i2} \end{bmatrix}$, according to (5).

After obtaining the proper values of Q_{ij}'s from the curves of Fig. 3 or Table I, the calculation of the four elements of the over-all Q matrix is straightforward. Thus, referring to (17),

$$\begin{aligned} Q_1 &= Q_{11} + \gamma_1 Q'_{21} + (1-\gamma_1)Q''_{21} \\ Q_2 &= Q_{11} + \gamma_1 Q'_{22} + (1-\gamma_1)Q''_{22} \\ Q_3 &= Q_{12} + (1-\partial_1)Q'_{21} + \partial_1 Q''_{21} \\ Q_4 &= Q_{12} + (1-\partial_1)Q'_{22} + \partial_1 Q''_{22}. \end{aligned} \quad (18)$$

Finally, the channel capacity is given, analogous to (6), by

$$C = \log \sum_{i=1}^{4} 2^{Q_i}$$

$$= \log (2^{Q_1} + 2^{Q_2} + 2^{Q_3} + 2^{Q_4}) \text{ bits/doublet.} \quad (19)$$

We shall illustrate the steps by a numerical example and discuss (19) under a few special conditions.

1) Assume, for a certain channel transmitting doublets,

$$\begin{aligned} \alpha_1 &= 0.80 \qquad \beta_1 = 0.90 \\ \alpha'_2 &= 0.70 \qquad \beta'_2 = 0.75 \\ \alpha''_2 &= 0.50 \qquad \beta''_2 = 0.60. \end{aligned} \quad (20)$$

From Table I, the following Q_{ij}'s are read:

$$\begin{aligned} Q_{11} &= -0.794194 \qquad Q_{12} = -0.432861 \\ Q'_{21} &= -0.927966 \qquad Q'_{22} = -0.772381 \\ Q''_{21} &= -1.145250 \qquad Q''_{22} = -0.854750. \end{aligned} \quad (21)$$

From (18),

$$\begin{aligned} Q_1 &= -1.660079 \qquad Q_2 = -1.543040 \\ Q_3 &= -1.609151 \qquad Q_4 = -1.299364. \end{aligned} \quad (22)$$

The resultant capacity is

$$C = 0.478922 \text{ bits/doublet.} \quad (23)$$

2) If the first digit is errorless in transmission, *i.e.*, $\alpha_1 = \beta_1 = 1$, then $\gamma_1 = \partial_1 = 1$, $Q_{11} = Q_{12} = 0$. From (17),

$$Q = \begin{bmatrix} Q'_{21} \\ Q'_{22} \\ Q''_{21} \\ Q''_{22} \end{bmatrix} \quad (24)$$

and

$$C = \log (2^{C_2'} + 2^{C_2''}). \quad (25)$$

The case is equivalent to two ABC (wnm) in parallel. The capacity of each channel C'_2 and C''_2 is determined by the noise probability matrices of the second digit, A'_2 and A''_2, respectively. This point will be clearer if one refers to the matrix A of (12). Here the submatrices of the upper right and lower left corners are replaced by 0, and the doublets 00, 01, and 10, 11 supply in an equivalent manner two separate binary channels with no interconnection. It is also evident that the same result will be obtained if $\alpha_1 = \beta_1 = 0$.

3) If the transmission property of the second digit is independent of the first digit, although not necessarily equal to that of the first digit, then

$$A'_2 = A''_2 = A_2 \quad (26)$$

and

$$Q = \begin{bmatrix} Q_{11} + Q_{21} \\ Q_{11} + Q_{22} \\ Q_{12} + Q_{21} \\ Q_{12} + Q_{22} \end{bmatrix}. \quad (27)$$

It may be easily shown that, in this case,

$$C = C_1 + C_2 \quad (28)$$

where C_1 and C_2 are capacities of ABC (wnm) corresponding to A_1 and A_2, respectively. If, further, $A_1 = A_2$, then the channel is memory-less and

$$C = 2C_1 \text{ bits/doublet,} \quad (29)$$

$$\text{or } C = C_1 \text{ bits/digit,}$$

as it should be.

4) The limiting conditions under which the channel becomes either perfect or capacity-less can be seen more easily by examining first the case of the ABC (wnm). The perfect capacity ($C = 1$ bit/digit) is obtained when either $\alpha = \beta = 1$, or $\alpha = \beta = 0$. The channel possesses zero capacity when $\alpha + \beta = 1$, or the 2×2 determinant of the single digit matrix $| A | = 0$.

The same conditions apply to the doublets, except that all three submatrices A_1, A_2', and A_2'' must satisfy the same or complementary conditions, for example, as in the following.

Perfect Channel:

$$A_1 = A_2' = A_2'' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

or

$$A_1 = A_2 = A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{30}$$

or any combinations of A_1, A_2', and A_2'' which satisfy either of the two conditions.

Zero-Capacity Channel:

The doublet matrix has identical rows and its 4×4 determinant

$$| A | = 0. \tag{31}$$

Extension to the Triplet

The generalization from doublet to triplet transmission is relatively simple. Four noise probability matrices A_3', A_3'', A_3''', and A_3'''' should be added to represent the property of the channel transmission for the third digit following the four possible doublets. The over-all noise probability matrix for the triplet can be composed by the pyramid structure of these matrices and the previously defined matrices in a way similar to (13).

$$A = \begin{bmatrix} A_3' & & & \\ & A_3'' & & \\ & & A_3''' & \\ & & & A_3'''' \end{bmatrix} \begin{bmatrix} A_2' & \\ & A_2'' \end{bmatrix} A_1. \tag{32}$$

During the multiplication process, to be carried from the right to the left, the submatrices in the left should be first considered merely as elements and then expanded.

Through postmultiplication of the inverted matrix with the negative of the noise entropy matrix, the auxiliary matrix Q is obtained again. The latter is expanded in a detailed manner which is directly applicable for the computation of C. This is shown as (33).

$$\begin{aligned} Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \\ Q_7 \\ Q_8 \end{bmatrix} = \Bigg\{ & \begin{bmatrix} Q_{11} \\ Q_{11} \\ Q_{11} \\ Q_{11} \end{bmatrix} + \gamma_1 \begin{bmatrix} Q_{21}' \\ Q_{21}' \\ Q_{22}' \\ Q_{22}' \end{bmatrix} + (1 - \gamma_1) \begin{bmatrix} Q_{21}'' \\ Q_{21}'' \\ Q_{22}'' \\ Q_{22}'' \end{bmatrix} + \gamma_1 \begin{bmatrix} \gamma_2' \begin{bmatrix} Q_{31}' \\ Q_{32}' \end{bmatrix} + (1 - \gamma_2') \begin{bmatrix} Q_{31}'' \\ Q_{32}'' \end{bmatrix} \\ (1 - \delta_2') \begin{bmatrix} Q_{31}' \\ Q_{32}' \end{bmatrix} + \delta_2' \begin{bmatrix} Q_{31}'' \\ Q_{32}'' \end{bmatrix} \end{bmatrix} \\ & \begin{bmatrix} Q_{12} \\ Q_{12} \\ Q_{12} \\ Q_{12} \end{bmatrix} + (1 - \delta_1) \begin{bmatrix} Q_{21}' \\ Q_{21}' \\ Q_{22}' \\ Q_{22}' \end{bmatrix} + \delta_1 \begin{bmatrix} Q_{21}'' \\ Q_{21}'' \\ Q_{22}'' \\ Q_{22}'' \end{bmatrix} + (1 - \delta_1) \begin{bmatrix} \gamma_2' \begin{bmatrix} Q_{31}' \\ Q_{32}' \end{bmatrix} + (1 - \gamma_2') \begin{bmatrix} Q_{31}'' \\ Q_{32}'' \end{bmatrix} \\ (1 - \delta_2') \begin{bmatrix} Q_{31}' \\ Q_{32}' \end{bmatrix} + \delta_2' \begin{bmatrix} Q_{31}'' \\ Q_{32}'' \end{bmatrix} \end{bmatrix} \\ & + (1 - \gamma_1) \begin{bmatrix} \gamma_2'' \begin{bmatrix} Q_{31}''' \\ Q_{32}''' \end{bmatrix} + (1 - \gamma_2'') \begin{bmatrix} Q_{31}'''' \\ Q_{32}'''' \end{bmatrix} \\ (1 - \delta_2'') \begin{bmatrix} Q_{31}''' \\ Q_{32}''' \end{bmatrix} + \delta_2'' \begin{bmatrix} Q_{31}'''' \\ Q_{32}'''' \end{bmatrix} \end{bmatrix} \\ & + \delta_1 \begin{bmatrix} \gamma_2'' \begin{bmatrix} Q_{31}''' \\ Q_{32}''' \end{bmatrix} + (1 - \gamma_2'') \begin{bmatrix} Q_{31}'''' \\ Q_{32}'''' \end{bmatrix} \\ (1 - \delta_2'') \begin{bmatrix} Q_{31}''' \\ Q_{32}''' \end{bmatrix} + \delta_2'' \begin{bmatrix} Q_{31}'''' \\ Q_{32}'''' \end{bmatrix} \end{bmatrix} \Bigg\}. \end{aligned} \tag{33}$$

Despite the formidable appearance of this matrix it can be set up by simple rules as shown in the next section. Most important, all the Q_{ij}'s can be read from Fig. 3 or Table I. This results in a considerable reduction in computation as opposed to 8×8 matrix operations or maximization processes with several constraints.

Iterative Procedure for the N-Plet

To extend the method to one which is applicable to blocks of n digits, it is desirable to adopt an iterative procedure. The procedure is best described by letting q_1, $q_2, q_3, \cdots, q_n$ be the notation of Q for single digits, doublets, triplets $\cdots$ n plets, respectively. Superscripts, 0 or 1, will be used to differentiate whether the single digit, doublet, etc., is preceded by "0" or "1." Thus q_1^0 is the Q of a single digit preceded by "0," and q_2^1 is the Q of a doublet preceded by "1." Also, the notation for the matrix $\begin{bmatrix} Q_{11} \\ Q_{11} \end{bmatrix}$ where Q_{11} is used twice will be simplified to read $[Q_{11}]$ (2). The same applies to the matrices of Q_{12}, and other parenthetical subscripts representing the number of iterations.

With this notation at hand, it can be shown that the equations for $q_1, q_2, q_3, \cdots, q_n$ can be expressed as

$$q_1 = \begin{bmatrix} Q_{11} \\ Q_{12} \end{bmatrix} \tag{34a}$$

$$q_2 = \begin{bmatrix} [Q_{11}](2) \\ [Q_{12}](2) \end{bmatrix} + A_1^{-1} \begin{bmatrix} q_1^0 \\ q_1^1 \end{bmatrix} \tag{34b}$$

$$q_3 = \begin{bmatrix} [Q_{11}](4) \\ [Q_{12}](4) \end{bmatrix} + A_1^{-1} \begin{bmatrix} q_2^0 \\ q_2^1 \end{bmatrix} \tag{34c}$$

$$\vdots \qquad\qquad \vdots$$

$$q_n = \begin{bmatrix} [Q_{11}](2^{n-1}) \\ [Q_{12}](2^{n-1}) \end{bmatrix} + A_1^{-1} \begin{bmatrix} q_{n-1}^0 \\ q_{n-1}^1 \end{bmatrix}. \tag{34d}$$

Thus the calculation of q_n must be preceded by the determination of the following number of q's:

2^{n-1}	q_1's
2^{n-2}	q_2's
2^{n-3}	q_3's
.....	
2	q_{n-1}'s.

The last two q_{n-1} matrices are then substituted into (34d) to obtain q_n. The latter is a column matrix with 2^n rows.

It can be easily verified, for example, that this procedure will yield the form of q_2 as given by (17) for the doublet and that of q_3 as given by (33) for the triplet.

Properties similar to those discussed under the section on the doublet can also be extended and proved for the general case of the n plet.

Probabilities for the Received and Transmitted Symbols

The channel capacity can be fully realized only if the symbols, here referred to as the n plets, are transmitted and received with appropriate probabilities. According to Muroga,[7] the probabilities of the received symbols for the capacity realization are given by the following column matrix:

$$P(y)] = 2^{-C}2^{Q}]. \tag{35}$$

It is evident that the elements of $P(y)$ are nonnegative and add up to unity.

Recalling the fact that the noise probability matrix A contains as elements the conditional probabilities $p(y_i \mid x_i)$, the probabilities of the transmitted symbol corresponding to (34) are given by the row matrix

$$\underbrace{P(x)}_{} = \underbrace{P(y)}_{} A^{-1}. \tag{36}$$

A numerical example is given by assigning to α's and β's the same values as in (20). The use of (35) and (36) gives the results:

$$P(y)] = \begin{bmatrix} 0.227043 \\ 0.246224 \\ 0,235200 \\ 0.291533 \end{bmatrix} \quad \text{and} \quad P(x)] = \begin{bmatrix} 0.277784 \\ 0.255454 \\ 0.172266 \\ 0.294495 \end{bmatrix} \tag{37}$$

where the four rows are arranged according to the order of the doublets 00, 01, 10, and 11, respectively. The sum of the elements of $P(y)$ and that of $P(x)$ are both equal to unity.

As pointed out by Muroga, some of the probabilities of the transmitted symbols, in certain conditions not explicitly known, may turn out to be negative and thus are not physically possible. It then is necessary to suppress some of the transmitted symbols for capacity calculation and realization. However, in the ABC (wnm) it has been demonstrated by Silverman[8] that these required probabilities are always positive and lie in the range $1/e$ to $1 - 1/e$. In the case of doublets, if $A_2' = A_2'' = A_2$ it can be shown that this property is again satisfied by each digit. Actually the capacity under this condition is given by $C = C_1 + C_2$ as derived previously. The two digits in such a doublet are transmitted independently and the required probability for each doublet can be shown to be the product of the probabilities of the first and second digit when coded to realize the capacities C_1 and C_2, respectively. There is no guarantee of the nonnegativeness of the elements of $P(x)$ in the general case of ABC (wfm). A more complicated procedure of capacity computation will have to be used under adverse conditions.

Discussion

It has been shown in this paper that when equal blocks of binary digits are used to transmit or store information through an asymmetrical channel with finite memory, the capacity of this channel can be calculated relatively easily by the use of curves or a table. The need for the consideration of channel memory arises from such phenomena as multipath transmission and interdigit interference due to limited bandwidth-time product of the transmitting and receiving equipment and of the transmission medium. It may arise also when the successive digits are intentionally assigned with different weights (weighted PCM). Since the channel capacity is the least

upper bound of the rate of transmission, it serves as an important criterion by means of which the performance of coding, detection and noise reduction schemes can be measured. The determination of the noise probabilities required in the capacity calculation of various channels forms an important subject in itself. The design of appropriate coding schemes to match with the asymmetrical channel with finite memory will be an intriguing problem for mathematicians and communication engineers.

Some calculations have been made recently on a hypothetical channel using the present method, which reveal certain interesting relations between the effects of the signal-to-noise ratio and the bandwidth (upon which the memory depends) on the capacity of the channel. It is hoped to publish these results later.

ACKNOWLEDGMENT

The paper by Muroga was first brought to the attention of the author and made available to him by Prof. R. M. Fano of M.I.T.

The numerical calculations on which Fig. 3, Fig. 4, and Table I are based were done by P. G. McHugh and Jerry DiPalma.

The author also acknowledges the help of R. D. Klein with the English construction in several paragraphs of this paper.

The Fourth Product Moment of Infinitely Clipped Noise*

J. A. McFADDEN†

Summary—**This paper considers the fourth product moment, $w(\tau_1, \tau_2, \tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)]$, when $x(t)$ is infinitely clipped noise with a mean value of zero. If the noise is Gaussian before clipping, the moment w is not obtainable in closed form. For this reason, the Gaussian assumption is withdrawn and other assumptions are employed. If the zeros of $x(t)$ obey the Poisson distribution, a particularly simple result follows for w and for all higher moments. An alternative assumption is the following. Let unspecified events occur at times t_0, t_1, t_2, $\cdots$ according to the Poisson distribution. If alternate events, *i.e.*, those at t_1, t_3, t_5, $\cdots$, are designated as the zeros of $x(t)$, both the autocorrelation function and $w(\tau_1, \tau_2, \tau_3)$ can be derived. The results are in terms of elementary functions. A comparison is made between these models and clipped Gaussian processes.**

INTRODUCTION

CONSIDER a random process $\xi(t)$ which is both stationary and ergodic. The mean value $E[\xi(t)]$ is zero and the process is symmetric about the mean.

Let $x(t)$ denote the output after the process $\xi(t)$ is infinitely clipped; *i.e.*, let

$$\begin{aligned} x(t) &= 1 \quad \text{if} \quad \xi(t) \geq 0; \\ &= -1 \quad \text{if} \quad \xi(t) < 0. \end{aligned} \tag{1}$$

Now if the input $\xi(t)$ is assumed to be Gaussian, the autocorrelation function of the output $x(t)$, namely $r(\tau) = E[x(t)x(t+\tau)]$, is derivable from the normalized autocorrelation function of $\xi(t)$ by a well-known arcsine formula.[1] On the other hand, for the fourth product moment,

$$w(\tau_1, \tau_2, \tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)], \tag{2}$$

no such simple formula is known.

The moment $w(\tau_1, \tau_2, \tau_3)$ is often needed in problems concerning clipped noise. For example, if the autocorrelation function $r(\tau)$ is to be estimated by means of a time average over a *finite time*, the variance of this estimator involves a double indefinite integral of $w(\tau_1, \tau_2, \tau_3)$ with respect to the arguments.[2]

The moment $w(\tau_1, \tau_2, \tau_3)$ is closely related to the quadrivariate normal integral, which is the probability that four random variables which are jointly Gaussian are simultaneously positive.[3,4] That integral has not been solved in closed form except for special numerical values of the correlation matrix. Special solutions are not usually sufficient, since the solution for $w(\tau_1, \tau_2, \tau_3)$ must be known for *all* values of the τ_i so that it can be integrated with respect to the τ_i.

For the same reason, numerical evaluation of the quadrivariate normal integral is not of much use in noise problems of the type mentioned above.

What is really needed is a simple expression containing easily integrable functions of τ_1, τ_2, and τ_3. Such a solution is apparently not available under the given assumptions.

* Manuscript received by the PGIT, June 2, 1958.

† School of Elec. Eng., Purdue University, Lafayette, Ind.

[1] See, for example, J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," McGraw-Hill Book Co., Inc., New York, N. Y., p. 58; 1950.

[2] J. H. Laning, Jr., and R. H. Battin, "Random Processes in Automatic Control," McGraw-Hill Book Co., Inc., New York, N. Y., p. 161; 1956.

[3] J. A. McFadden, "The axis-crossing intervals of random functions II," IRE TRANS. ON INFORMATION THEORY, vol. IT-4, pp. 14–24; March, 1958. See (84).

[4] J. A. McFadden, "Urn models of correlation and a comparison with the multivariate normal integral," *Ann. Math. Stat.*, vol. 26, pp. 478–489; September, 1955. See sec. 6 and bibliography.

Also, J. A. McFadden, "An approximation for the symmetric, quadrivariate normal integral," *Biometrika*, vol. 43, pp. 206–207; June, 1956.

[In fact, even the correlation function $r(\tau)$ is not a particularly simple function of τ, *i.e.*, one that can be integrated in closed form.]

What then can be done in a practical problem in which the moment $w(\tau_1, \tau_2, \tau_3)$ is needed? The method of this paper is to remove the assumption that $\xi(t)$ is Gaussian and to prescribe instead the properties of the axis crossings of $\xi(t)$ [which are also the axis crossings of $x(t)$].

Two different assumptions are discussed:

a) The zeros obey the Poisson distribution. This is one of Rice's "random telegraphic signals."[5]

b) A random sequence of events occurs at various instants according to the Poisson distribution, and alternate events are designated as the zeros of $x(t)$.

Derivation

The results for $w(\tau_1, \tau_2, \tau_3)$ are now derived.

Consider a random sequence of time intervals which end at the instants $t_0, t_1, t_2, \cdots$. The number of these instants occurring between the times t and $t + \tau$ obeys the Poisson distribution; that is, the probability that this number is exactly n is

$$h(n, \tau) = \frac{(\alpha\tau)^n}{n!} e^{-\alpha\tau}, \tag{3}$$

where α is the expected number of these instants occurring in a unit interval of time.

Let the function $x(t)$ assume the constant value a_n over the interval $t_{n-1} < t \leq t_n$, where $n = 1, 2, 3, \cdots$. Each a_n is a random variable which assumes the value $+1$ or -1 with probability one half. The successive values $a_n, a_{n+1}, a_{n+2}, \cdots$ are *not* independent.[6] The form of dependence is prescribed later.

Consider now the product $x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3)$, where $0 \leq \tau_1 \leq \tau_2 \leq \tau_3$. If t falls between t_{k-1} and t_k, then $x(t) = a_k$. Let the number of end-points of intervals between t and $t + \tau_1$ be l. Let the number between $t + \tau_1$ and $t + \tau_2$ be m, and let the number between $t + \tau_2$ and $t + \tau_3$ be n. Then the fourfold product is

$$x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3) = a_k a_{k+l} a_{k+l+m} a_{k+l+m+n}. \tag{4}$$

Then the expected value of this product is the product of the a's times the joint probability that exactly l, m, and n intervals end, respectively, between t and $t + \tau_1$, between $t + \tau_1$ and $t + \tau_2$, and between $t + \tau_2$ and $t + \tau_3$, summed over all the values of l, m, and n, and averaged over k. But since the instants t_i occur according to the Poisson distribution, it follows that the numbers l, m, and n are statistically independent. Then $w(\tau_1, \tau_2, \tau_3)$, the expected value of (4), is

$$w(\tau_1, \tau_2, \tau_3) = E\left\{ \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} s_{klmn} \times h(l, \tau_1)h(m, \tau_2 - \tau_1)h(n, \tau_3 - \tau_2) \right\}, \tag{5}$$

where E denotes the average over k and where

$$s_{klmn} = a_k a_{k+l} a_{k+l+m} a_{k+l+m+n}. \tag{6}$$

The next step is to specify s_{klmn} for particular models. Under assumption a), the zeros of $x(t)$ obey the Poisson distribution. This requirement can easily be imposed on (5) if it is prescribed that

$$a_{i+1} = -a_i; \tag{7}$$

that is, the sign of $x(t)$ must change at each of the instants $t_0, t_1, t_2, \cdots$. Then

$$\begin{aligned} a_{k+l} &= (-1)^l a_k, \\ a_{k+l+m} &= (-1)^{l+m} a_k, \\ a_{k+l+m+n} &= (-1)^{l+m+n} a_k; \end{aligned} \tag{8}$$

therefore

$$\begin{aligned} s_{klmn} &= (-1)^{3l+2m+n} a_k^4 \\ &= (-1)^{l+n}. \end{aligned} \tag{9}$$

Since this result is independent of k, it need not be averaged over k. Then the triple sum (5) is easily evaluated; the result is[7]

$$w(\tau_1, \tau_2, \tau_3) = e^{-2\alpha\tau_1} e^{-2\alpha(\tau_3 - \tau_2)}. \tag{10}$$

Note that $w(\tau_1, \tau_2, \tau_3)$ is completely specified if the first and third time differences, τ_1 and $(\tau_3 - \tau_2)$, are known; the second one, $(\tau_2 - \tau_1)$, is not needed.

Now for this model the autocorrelation function is[5]

$$r(\tau) = e^{-2\alpha|\tau|}; \tag{11}$$

therefore (10) can be written

$$w(\tau_1, \tau_2, \tau_3) = r(\tau_1)r(\tau_3 - \tau_2). \tag{12}$$

This result can be extended to the higher moments of even order. Let $x_1 = x(t)$ and let $x_{i+1} = x(t + \tau_i)$, where $i = 1, 2, 3, \cdots$ and $\tau_{i+1} \geq \tau_i$. It has been shown in (12) that

$$E(x_1x_2x_3x_4) = E(x_1x_2)E(x_3x_4). \tag{13}$$

It follows directly that the products x_1x_2 and x_3x_4 are uncorrelated. But since x_1x_2 and x_3x_4 can each assume only two values, $+1$ and -1, the products must be independent.[8] In this case it can be proved also that *all* the successive products $x_1x_2, x_3x_4, x_5x_6, \cdots$ are mutually independent; therefore[9]

[5] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–332; July, 1944, and vol. 24, pp. 46–156; January, 1945. See sec. 2.7.

[6] Cf. H. M. James, N. B. Nichols, and R. S. Phillips, "Theory of Servomechanisms," McGraw-Hill Book Co., Inc., New York, N. Y., p. 301; 1947, where the a_n are independent. The present method is a generalization of the derivation on pp. 301–303.

[7] Cf. T. A. Magness, "Spectral response of a quadratic device," *J. Appl. Phys.*, vol. 25, pp. 1357–1365; November, 1954. Eq. (57).

[8] W. Feller, "An Introduction to Probability Theory and Its Applications," John Wiley and Sons, Inc., New York, N. Y., vol. 1, 1st ed., p. 189; 1950. See problem 21.

[9] This result was first shown to the author by G. R. Cooper in his memorandum, "System measurement with binary noise," (to be published).

$$E(x_1x_2 \cdots x_{2n-1}x_{2n}) = E(x_1x_2) \cdots E(x_{2n-1}x_{2n}) = e^{-2\alpha\tau_1} \cdots e^{-2\alpha(\tau_{2n-1}-\tau_{2n-2})}. \quad (14)$$

The fourth moment is now derived under assumption b). In this case, alternate instants t_1, t_3, t_5, $\cdots$ are designated as the zeros of the process $x(t)$. In other words,

$$a_{2i+1} = a_{2i}, \qquad a_{2i+2} = -a_{2i+1}, \quad (15)$$

where $i = 1, 2, 3, \cdots$. Then there are three cases to be considered. The results are as follows.

1) If l and n are both even,

$$s_{klmn} = (-1)^{(l+n)/2}. \quad (16)$$

2) If l and n are both odd, then

$$s_{klmn} = (-1)^{(l+n)/2+m}. \quad (17)$$

3) If l is odd and n is even, then

$$s_{klmn} = (-1)^{k+(l-1)/2+n/2}. \quad (18)$$

But since k can be either even or odd, the average over k is zero. The same is true for the case in which l is even and n is odd.

Substitution of the above results into (5) yields the following:

$$\begin{aligned} w(\tau_1, \tau_2, \tau_3) &= \sum_{\substack{l \\ \text{even}}} \sum_{m} \sum_{\substack{n \\ \text{even}}} (-1)^{(l+n)/2} h(l, \tau_1)h(m, \tau_2 - \tau_1)h(n, \tau_3 - \tau_2) \\ &\quad + \sum_{\substack{l \\ \text{odd}}} \sum_{m} \sum_{\substack{n \\ \text{odd}}} (-1)^{(l+n)/2+m} \\ &\quad \times h(l, \tau_1)h(m, \tau_2 - \tau_1)h(n, \tau_3 - \tau_2) \\ &= e^{-\alpha[\tau_1+(\tau_3-\tau_2)]}[\cos\alpha\tau_1 \cos\alpha(\tau_3 - \tau_2) \\ &\quad - e^{-2\alpha(\tau_2-\tau_1)} \sin\alpha\tau_1 \sin\alpha(\tau_3 - \tau_2)]. \end{aligned} \quad (19)$$

The autocorrelation function of this process may be obtained as a special case. If $\tau_3 \to \tau_2$, then $w(\tau_1, \tau_2, \tau_3)$ becomes

$$E[x(t)x(t + \tau_1)x^2(t + \tau_2)] = E[x(t)x(t + \tau_1)].$$

With $\tau_3 = \tau_2$ and $\tau_1 = \tau$, (19) becomes

$$r(\tau) = e^{-\alpha|\tau|} \cos\alpha\tau. \quad (20)$$

The method used in obtaining (19) may be extended to the case of higher moments of even order, but with increasing complexity.

Discussion

Of what value are the above results when the actual problem involves infinitely clipped *Gaussian* noise?

Consider a Gaussian process $\xi(t)$ which has the normalized autocorrelation function {*i.e.*, $E[\xi(t)\xi(t + \tau)]$ divided by the variance},

$$\rho(\tau) = \sin\left[\frac{\pi}{2} r(\tau)\right], \quad (21)$$

where $r(\tau)$ is given by (11). (It is not clear whether such a correlation function is possible; it remains to be proved that the corresponding spectral density is a nonnegative function.) Provided such a Gaussian process exists, the autocorrelation function $r(\tau)$ [after the clipping process (1)] is identical with (11), because of the arcsine formula.[1]

However, the clipped Gaussian process cannot be identical with the Poisson model a). It can be shown,[10,11] for certain types of correlation functions, that the successive axis-crossing intervals of a Gaussian process cannot be statistically independent. This argument makes the Poisson model incompatible with the Gaussian assumption.

Evidently such a clipped Gaussian process and the Poisson model have the same autocorrelation function (11) but are not the same process; therefore, some of their higher-order moments must differ. It remains to be shown how well the fourth product moment (10) compares with the $w(\tau_1, \tau_2, \tau_3)$ for the corresponding clipped Gaussian process.

As stated earlier, $w(\tau_1, \tau_2, \tau_3)$ has not been obtained in closed form for a clipped Gaussian process. However, there is one limiting case which can be used for comparison. This limiting case occurs when $0 \leq \tau_1 \leq \tau_2 \leq \tau_3$ and the variables τ_1 and $\tau_3 - \tau_2$ become infinitesimal.

It has been shown,[12] for a symmetric process (not necessarily Gaussian), that the limiting form of the fourth product moment after clipping is simply related to the function $U(\tau)$, where $U(\tau)d\tau$ is the conditional probability that a zero will occur between $t + \tau$ and $t + \tau + d\tau$, given a zero at time t. The relation is the following:

$$U(\tau) = \frac{1}{4\beta}\left(\frac{\partial^2 w}{\partial\tau_1\,\partial\tau_3}\right)_{0+,\tau,\tau+} \quad (22)$$

where β is the expected number of axis crossings per unit time.

Now $U(\tau)$ is known both for the Poisson model a) and for a Gaussian process. For the Poisson model, the moment (10) may be substituted into (22), with $\beta = \alpha$, and the result is a constant.

$$U(\tau) = \alpha \quad (23)$$

for all $\tau > 0$. This result checks the basic Poisson assumption, namely that the probability of a zero occurring between t and $t + dt$ is αdt, independent of the location of all previous zeros.

For a Gaussian process $U(\tau)d\tau$ has been calculated by Rice.[13] If the autocorrelation function (21) is substituted into Rice's formula, using (11) for $r(\tau)$, the result is some-

[10] McFadden, footnote 3, p. 17.
[11] D. S. Palmer, "Properties of random functions," *Proc. Camb. Phil. Soc.*, vol. 52, pp. 672–686; October, 1956. See pp. 679–680.
[12] McFadden, footnote 3, (11).
[13] Rice, *op. cit.*, (3.4–10).

what different from (23). $U(\tau)$ is a monotonically decreasing function, beginning[14] with $U(0) = 1.436\alpha$ and approaching the value α as $\tau \to \infty$. The decay is roughly exponential, with a time constant close to $1/(4\alpha)$, *i.e.*, one quarter of the average time between axis crossings.

It may be concluded that the Poisson model a) is not a very good approximation to a clipped Gaussian process; yet (10) is still recommended for the fourth product moment because of its simplicity and the absence of more suitable models. It should be tried whenever the autocorrelation function $\rho(\tau)$ of the original Gaussian process $\xi(t)$ has a behavior somewhat like (21), where $r(\tau)$ is given by (11). α should be chosen equal to β, the expected number of zeros per unit time.[15]

A similar argument may be applied to the alternate Poisson model b). Consider a Gaussian process with an autocorrelation function (21), where now $r(\tau)$ is given by (20). Provided this process exists, it may be compared, after clipping, with the alternate Poisson model.

As before, the two processes cannot be identical. Although the previous arguments[10,11] do not apply to this particular $\rho(\tau)$, a slight extension shows that the successive axis-crossing intervals of this Gaussian process must also be statistically dependent. This argument makes the alternate Poisson model incompatible with the Gaussian assumption.

$U(\tau)$ may be calculated for the alternate Poisson model by substituting the moment (19) into (22) with $\beta = \alpha/2$. The result is the following:[16]

$$U(\tau) = \frac{\alpha}{2}(1 - e^{-2\alpha\tau}) \tag{24}$$

for all $\tau > 0$.

For the corresponding Gaussian process, the function $U(\tau)$ is quite similar to (24). It begins at $\tau = 0$ with exactly the same initial value, slope, and curvature as (24), then rises slightly above (24), and approaches the asymptotic value $\alpha/2$ while undergoing small oscillations.

The close agreement of $U(\tau)$ provides only a partial check on (19) for the fourth product moment. Again, however, the result is recommended because of its simplicity. It should be used when the autocorrelation function $\rho(\tau)$ of the original Gaussian process has a behavior somewhat like (21), where $r(\tau)$ is given by (20). α should be chosen equal to 2β, where β is the expected number of zeros per unit time.

An interesting problem for further research would be to devise other comparisons which might prove how closely the alternate Poisson model b) approximates a clipped Gaussian process.

One method of comparison is suggested by some recent work of Longuet-Higgins,[17] who has obtained certain derivatives of the quadrivariate normal integral. The relation to the fourth product moment after clipping has been described previously.[3] In terms of $w(\tau_1, \tau_2, \tau_3)$, Longuet-Higgins' results would provide the following.

Let the observations on $x(t)$ be performed at equal intervals. That is, let $\tau_1 = \frac{1}{3}\tau_3$ and $\tau_2 = \frac{2}{3}\tau_3$. Furthermore, let $t_1 = t$, $t_2 = t + \tau_1$, $t_3 = t + \tau_2$, and $t_4 = t + \tau_3$. Under these conditions, the results would provide expressions for

$$\frac{\partial^2 w}{\partial \tau_3^2} \quad \text{and} \quad \frac{\partial^2 w}{\partial t_1 \, \partial t_4}$$

in terms of the autocorrelation function $\rho(\tau)$.

Unfortunately these results are extremely complicated; yet if they were obtained numerically for the cases in question they would provide another comparison of clipped Gaussian noise with the models a) and b).

Another possibility for a comparison would be to evaluate the quadrivariate normal integral numerically for certain values of τ_1, τ_2, and τ_3, using the method given by Plackett.[18]

[14] McFadden, footnote 3, (24).
[15] Rice, *op. cit.*, (3.3-11).
[16] Cf. M. S. Bartlett, "An Introduction to Stochastic Process," Cambridge University Press, Cambridge, Eng., p. 167; 1955.
[17] M. S. Longuet-Higgins, "On the intervals between sucessive zeros of a random function," *Proc. Roy. Soc., London*, vol. A246, pp. 99–118; July, 1958.
[18] R. L. Plackett, "A reduction formula for multivariate normal integrals," *Biometrika*, vol. 41, pp. 351–360; December, 1954.

A Markoff Envelope Process*

JOHN N. PIERCE†

Summary—It is shown that the envelope of a narrow-band Gaussian noise constitutes a first-order Markoff process if the power spectrum of the noise is the same as would be obtained from a singly tuned RLC filter with white noise at the input.

Introduction

In his famous paper on random noise, Rice[1] derived the joint probability density function for two samples from the envelope of a narrow-band Gaussian noise. The problem of extending the solution to N samples of the envelope is very difficult. Kac and Siegert[2] and Hoffman[3] gave general solutions, although not in closed form. Since these solutions place no restrictions on the power spectrum of the noise, the results are intractable to most further manipulations.

If we restrict our attention to a power spectrum of the form

$$w(f) = \frac{1}{(f - f_m)^2 + F^2},$$

with f the frequency, f_m the center frequency, and F proportional to bandwidth, it is possible to arrive at a simple joint density function for N samples of the envelope. Such a spectrum is approximately that of the output of a high-Q singly tuned RLC filter with white Gaussian noise at the input.

In addition, it can be shown that the envelope of noise with this power spectrum constitutes a first-order Markoff process and is completely specified by the joint density function for two samples. The author shows this by the same method as used by Wang and Uhlenbeck,[4] that is, by showing that the conditional density function of the Nth envelope sample, given the $(N - 1)$ preceding envelope samples, is identical with the conditional density function given the immediately preceding sample. To reduce notational difficulties, however, a full derivation is given for only three envelope samples, with modifications of the derivation to extend it to N samples. In the following, the term "Markoff" will be taken to mean first-order Markoff.

* Manuscript received by the PGIT, July 16, 1958; revised manuscript received, August 6, 1958.

† Air Res. and Dev. Command, AF Cambridge Res. Center, Bedford, Mass.

[1] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 75–79; January, 1945.

[2] M. Kac and A. J. F. Siegert, "On the theory of noise in radio receivers with square law detectors," *J. Appl. Phys.*, vol. 18, pp. 396–397; April, 1947.

[3] W. C. Hoffman, "The joint distributions of n successive outputs of a linear detector," *J. Appl. Phys.*, vol. 25, pp. 1006–1007; August, 1954.

[4] M. C. Wang and G. E. Uhlenbeck, "On the theory of Brownian motion II," *Rev. Mod. Phys.*, vol. 17, pp. 323–342; April-July, 1945.

Joint Density Function for Three Samples

Adhering closely to Rice's notation in the following derivation, let $w(f)$ be the power spectrum of a narrow-band Gaussian noise, and let f_m be a representative midband frequency. Consider the three functions

$$\begin{aligned} I_1(t) &= I(t), \\ I_2(t) &= I(t + \tau_{1,2}), \\ I_3(t) &= I(t + \tau_{1,2} + \tau_{2,3}), \end{aligned} \tag{1}$$

where $I(t)$ is the time function having the given power spectrum, $\tau_{1,2}$ is the delay between the first and second functions, and $\tau_{2,3}$ is the delay between the second and third. We can represent each function as being the sum of cosine and sine (in-phase and quadrature) components at the midband frequency:

$$\begin{aligned} I_1 &= I_{C1} \cos \omega_m t - I_{S1} \sin \omega_m t, \\ I_2 &= I_{C2} \cos \omega_m t - I_{S2} \sin \omega_m t, \\ I_3 &= I_{C3} \cos \omega_m t - I_{S3} \sin \omega_m t, \end{aligned} \tag{2}$$

where $\omega_m = 2\pi f_m$, and $I_{C1}, \cdots, I_{S3}$ vary at a rate slow compared with the midband frequency. For each of the three functions, we define the envelope R of the narrow-band function by

$$\begin{aligned} R &= (I_C^2 + I_S^2)^{1/2}, \\ I_C &= R \cos \theta, \\ I_S &= R \sin \theta. \end{aligned} \tag{3}$$

The six components, which we order in a row matrix,

$$\underline{X} = \underline{I_{C1}, I_{C2}, I_{C3}, I_{S1}, I_{S2}, I_{S3}}, \tag{4}$$

constitute a six-dimensional normal process with a moment matrix,

$$[M] = \begin{bmatrix} \psi_0 & \mu_{1,2} & \mu_{1,3} & 0 & \nu_{1,2} & \nu_{1,3} \\ \mu_{1,2} & \psi_0 & \mu_{2,3} & -\nu_{1,2} & 0 & \nu_{2,3} \\ \mu_{1,3} & \mu_{2,3} & \psi_0 & -\nu_{1,3} & -\nu_{2,3} & 0 \\ 0 & -\nu_{1,2} & -\nu_{1,3} & \psi_0 & \mu_{1,2} & \mu_{1,3} \\ \nu_{1,2} & 0 & -\nu_{2,3} & \mu_{1,2} & \psi_0 & \mu_{2,3} \\ \nu_{1,3} & \nu_{2,3} & 0 & \mu_{1,3} & \mu_{2,3} & \psi_0 \end{bmatrix}. \tag{5}$$

From Rice's equation pair (3.7-11) and its preceding material, the coefficients in the moment matrix are given by

$$\psi_0 = \int_0^\infty w(f)\, df,$$

$$\mu_{1,2} = \int_0^\infty w(f) \cos [2\pi(f - f_m)\tau_{1,2}]\, df,$$

$$\mu_{2,3} = \int_0^\infty w(f) \cos [2\pi(f - f_m)\tau_{2,3}]\, df,$$

$$\mu_{1,3} = \int_0^\infty w(f) \cos [2\pi(f - f_m)(\tau_{1,2} + \tau_{2,3})]\, df, \qquad (6)$$

$$\nu_{1,2} = \int_0^\infty w(f) \sin [2\pi(f - f_m)\tau_{1,2}]\, df,$$

$$\nu_{2,3} = \int_0^\infty w(f) \sin [2\pi(f - f_m)\tau_{2,3}]\, df, \quad \text{and}$$

$$\nu_{1,3} = \int_0^\infty w(f) \sin [2\pi(f - f_m)(\tau_{1,2} + \tau_{2,3})]\, df.$$

We now choose the power spectrum

$$w(f) = \frac{k}{\alpha^2 + 16\pi^2(f - f_m)^2}, \qquad (7)$$

where α is the half-power bandwidth in radians per second, and k is an arbitrary constant. Then

$$\psi_0 = \int_0^\infty \frac{k\, df}{\alpha^2 + 16\pi^2(f - f_m)^2} = \int_{-f_m}^\infty \frac{k\, df}{\alpha^2 + 16\pi^2 f^2} = \frac{k}{4\pi\alpha} \int_{-2\pi f_m/\alpha}^\infty \frac{dx}{1 + x^2},$$

$$\mu = \int_0^\infty \frac{k \cos [2\pi(f - f_m)\tau]\, df}{\alpha^2 + 16\pi^2(f - f_m)^2} = \frac{k}{4\pi\alpha} \int_{2\pi f_m/\alpha}^\infty \frac{\cos [(\alpha/2)\tau x]\, dx}{1 + x^2},$$

$$\nu = \int_0^\infty \frac{k \sin [2\pi(f - f_m)\tau]\, df}{\alpha^2 + 16\pi^2(f - f_m)^2} = \frac{k}{4\pi\alpha} \int_{-2\pi f_m/\alpha}^\infty \frac{\sin [(\alpha/2)\tau x]\, dx}{1 + x^2}.$$

As $2\pi f_m/\alpha$ (the center frequency divided by the bandwidth) becomes large,

$$\psi_0 \to \frac{k}{4\alpha}, \qquad \mu \to \frac{k}{4\alpha} \exp(-\alpha\tau/2), \qquad \nu \to 0. \qquad (8)$$

For simplicity in the final result, one writes

$$b_{1,2} = \exp(-\alpha\tau_{1,2}), \qquad b_{2,3} = \exp(-\alpha\tau_{2,3}), \qquad (9)$$

so that

$$\mu_{1,2} = \psi_0\sqrt{b_{1,2}}, \quad \mu_{2,3} = \psi_0\sqrt{b_{2,3}}, \quad \mu_{1,3} = \psi_0\sqrt{b_{1,2}b_{2,3}}.$$

Since all the νs are zero, one is led to the obvious partitioning of the moment matrix for conciseness:

$$[M] = \psi_0 \left[\begin{array}{c|c} A & 0 \\ \hline 0 & A \end{array}\right] \qquad (10a)$$

where

$$[A] = \begin{bmatrix} 1 & \sqrt{b_{1,2}} & \sqrt{b_{1,2}b_{2,3}} \\ \sqrt{b_{1,2}} & 1 & \sqrt{b_{2,3}} \\ \sqrt{b_{1,2}b_{2,3}} & \sqrt{b_{2,3}} & 1 \end{bmatrix} \qquad (10b)$$

and, of course, the zero matrix is also 3×3.

The inverse and determinant of the moment matrix are also needed for the determination of the joint density function. These are given by

$$[M^{-1}] = \frac{1}{\psi_0} \left[\begin{array}{c|c} A^{-1} & 0 \\ \hline 0 & A^{-1} \end{array}\right], \qquad |M| = \psi_0^6 |A|^2, \qquad (11a)$$

where

$$[A^{-1}] = \begin{bmatrix} \dfrac{1}{(1 - b_{1,2})} & \dfrac{-\sqrt{b_{1,2}}}{(1 - b_{1,2})} & 0 \\ \dfrac{-\sqrt{b_{1,2}}}{(1 - b_{1,2})} & \dfrac{1 - b_{1,2}b_{2,3}}{(1 - b_{1,2})(1 - b_{2,3})} & \dfrac{-\sqrt{b_{2,3}}}{(1 - b_{2,3})} \\ 0 & \dfrac{-\sqrt{b_{2,3}}}{(1 - b_{2,3})} & \dfrac{1}{(1 - b_{2,3})} \end{bmatrix} \qquad (11b)$$

and

$$|A| = (1 - b_{1,2})(1 - b_{2,3}). \qquad (11c)$$

It should be pointed out at this juncture that the simple solution obtainable for the joint density function of the envelope samples is a result of the zero elements in the upper right and lower left corners of the inverse matrix A^{-1}. More precisely, the simple solution requires that only three diagonals contain nonzero elements.

With the row matrix $\underline{X}$ from (4) representing the six variates, the six-dimensional normal density function for the variates is

$$p(X) = \frac{1}{(2\pi)^3\sqrt{|M|}} \exp(-\tfrac{1}{2}X[M^{-1}]X_t), \qquad (12)$$

where $\underline{X}_t$ is the transpose, or column, matrix. Rather than write out this density function, one immediately makes the substitution indicated by (3), $I_C = R\cos\theta$, $I_S = R\sin\theta$. The differential element goes into $RdRd\theta$. After suitable trigonometric manipulations, the joint density function of the three Rs and three θs becomes

$$p(R_1, \theta_1, R_2, \theta_2, R_3, \theta_3) = \frac{R_1R_2R_3}{(2\pi)^3(1 - b_{1,2})(1 - b_{2,3})\psi_0^3} \cdot \exp\left[-\frac{R_1^2}{2(1 - b_{1,2})\psi_0} - \frac{(1 - b_{1,2}b_{2,3})R_2^2}{2(1 - b_{1,2})(1 - b_{2,3})\psi_0} - \frac{R_3^2}{2(1 - b_{2,3})\psi_0}\right] \exp\left[\frac{\sqrt{b_{1,2}}R_1R_2\cos(\theta_1 - \theta_2)}{(1 - b_{1,2})\psi_0} + \frac{\sqrt{b_{2,3}}R_2R_3\cos(\theta_2 - \theta_3)}{(1 - b_{2,3})\psi_0}\right]. \qquad (13)$$

Now,

$$p(R_1, R_2, R_3) = \iiint_0^{2\pi} p(R_1, \theta_1, R_2, \theta_2, R_3, \theta_3)\, d\theta_1\, d\theta_2\, d\theta_3. \quad (14)$$

If one temporarily replaces the coefficients of $\cos(\theta_1 - \theta_2)$ and $\cos(\theta_2 - \theta_3)$, in the exponent, by $d_{1,2}$ and $d_{2,3}$, respectively, the pertinent part of the integration is then

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta_3 \frac{1}{2\pi}\int_0^{2\pi} d\theta_2 \exp[d_{2,3}\cos(\theta_2 - \theta_3)]\frac{1}{2\pi} \cdot \int_0^{2\pi} d\theta_1 \exp[d_{1,2}\cos(\theta_1 - \theta_2)]$$

$$= \frac{1}{2\pi}\int_0^{2\pi} d\theta_3 \frac{1}{2\pi}\int_{-\theta_3}^{2\pi-\theta_3} dx \exp[d_{2,3}\cos x]\frac{1}{2\pi} \cdot \int_{-\theta_3-x}^{2\pi-\theta_3-x} dy \exp d_{1,2}\cos y,$$

x and y being dummy variables of integration. The periodicity of the two inner integrands permits one to change the limits of integration to $(0, 2\pi)$, so that the expression becomes

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta_3 \frac{1}{2\pi}\int_0^{2\pi} dx \exp[d_{2,3}\cos x]\frac{1}{2\pi} \cdot \int_0^{2\pi} dy \exp[d_{1,2}\cos y]$$

$$= I_0(d_{2,3})I_0(d_{1,2}).$$

The 3-sample density function is, therefore,

$$p(R_1, R_2, R_3) = \frac{R_1R_2R_3}{(1-b_{1,2})(1-b_{2,3})\psi_0^3} I_0\left(\frac{\sqrt{b_{1,2}}R_1R_2}{\psi_0(1-b_{1,2})}\right) I_0\left(\frac{\sqrt{b_{2,3}}R_2R_3}{\psi_0(1-b_{2,3})}\right) \cdot \exp\left[-\frac{R_1^2}{2(1-b_{1,2})\psi_0} - \frac{(1-b_{1,2}b_{2,3})R_2^2}{2(1-b_{1,2})(1-b_{2,3})\psi_0} - \frac{R_3^2}{2(1-b_{2,3})\psi_0}\right]. \quad (15)$$

Generalization to N Samples

Since the general method of solution for the N-sample case corresponds closely to that for three samples, only the points of divergence are given here. Eqs. (1)-(9) still apply after inclusion of all N variables in each equation. Again the matrices are written in the form

$$[M] = \psi_0\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}, \qquad [M^{-1}] = \frac{1}{\psi_0}\begin{bmatrix} C & 0 \\ 0 & C \end{bmatrix}$$

$$|M| = \psi_0^{2N}|A|^2,$$

where $C = A^{-1}$. For conciseness only the general terms of $[A]$ and $[C]$ are given:

$$a_{m,n} = \sqrt{(b_{m,m+1})(b_{m+1,m+2})\cdots(b_{n-1,n})} \text{ if } m < n,$$
$$= 1 \text{ if } m = n,$$
$$= a_{n,m} \text{ if } m > n; \quad (16)$$

with

$$b_{i,i+1} = \exp(-\tfrac{1}{2}\alpha\tau_{i,i+1}),$$

$$c_{m,n} = 0 \quad \text{if} \quad |m-n| > 1,$$

$$c_{m+1,m} = c_{m,m+1} = \frac{-\sqrt{b_{m,m+1}}}{1-b_{m,m+1}},$$

$$c_{11} = \frac{1}{1-b_{1,2}}, \qquad c_{NN} = \frac{1}{1-b_{N-1,N}}, \quad (17)$$

$$c_{m,m} = \frac{1-b_{m-1,m}b_{m,m+1}}{(1-b_{m-1,m})(1-b_{m,m+1})} \text{ if } m \neq 1 \text{ or } N.$$

$$A = \prod_{m=1}^{N-1}(1-b_{m,m+1}). \quad (18)$$

Substituting these matrices in (12) leads to a very involved $2N$- dimensional density function in the several Rs and θs. The multiple integral over the θs can fortunately be performed by the same technique as before—by substituting dummy variables and changing the limits of integration. The end result is the joint density function for N successive samples of the envelope:

$$p(R_1, R_2, \cdots, R_N) = \frac{R_1R_2\cdots R_N}{\psi_0^N}\prod_{m=1}^{N-1} I_0\left[\frac{\sqrt{b_{m,m+1}}R_mR_{m+1}}{\psi_0(1-b_{m,m+1})}\right] \cdot \exp\left[-\frac{R_1^2}{2(1-b_{1,2})\psi_0} - \frac{R_N^2}{2(1-b_{N-1,N})\psi_0} - \frac{1}{2\psi_0}\sum_{n=2}^{N-1}\frac{(1-b_{n-1,n}b_{n,n+1})R_n^2}{(1-b_{n-1,n})(1-b_{n,n+1})}\right]. \quad (19)$$

Conditional Density Function

The probability density of the Nth envelope sample, given the previous N-1 samples of the envelope, may be written as

$$p(R_N \mid R_1, R_2, \cdots, R_{N-1}) = \frac{p(R_1, R_2, \cdots, R_{N-1}, R_N)}{p(R_1, R_2, \cdots, R_{N-1})}, \quad (20)$$

or

$$= \frac{R_N}{(1-b_{N-1,N})\psi_0} \cdot \exp\left[-\frac{b_{N-1,N}R_{N-1}^2 + R_N^2}{2(1-b_{N-1,N})\psi_0}\right] I_0\left[\frac{\sqrt{b_{N-1,N}}R_{N-1}R_N}{(1-b_{N-1,N})\psi_0}\right]. \quad (21)$$

Eq. (21) was obtained by twice substituting (19) into (20). Rice's (3.7–10) and (3.7–13), modified to fit the notation used here, read

$$p(R_{N-1}) = \frac{R_{N-1}}{\psi_0} \exp\left(-\frac{R_{N-1}^2}{2\psi_0}\right),$$

$$p(R_{N-1}, R_N) = \frac{R_{N-1}R_N}{(1 - b_{N-1,N})\psi_0^2} \cdot \exp\left[-\frac{R_{N-1}^2 + R_N^2}{2(1 - b_{N-1,N})\psi_0}\right] I_0\left[\frac{\sqrt{b_{N-1,N}}R_{N-1}R_N}{(1 - b_{N-1,N})\psi_0}\right], \quad (22)$$

and since the conditional density of the Nth sample, given the $(N - 1)$th, is

$$p(R_N \mid R_{N-1}) = \frac{p(R_{N-1}, R_N)}{p(R_{N-1})},$$

we find by simple division that

$$p(R_N \mid R_{N-1}) = p(R_N \mid R_1, R_2, \cdots, R_{N-1}), \quad (23)$$

which is a necessary condition for the envelope to be a Markoff process. The other necessary conditions, as given in Wang and Uhlenbeck, are that

$$\left.\begin{aligned} p(R_N \mid R_{N-1}) &\geqq 0, \\ \int_0^\infty p(R_N \mid R_{N-1})\, dR_N &= 1, \\ p(R_N) &= \int_0^\infty p(R_N \mid R_{N-1}) p(R_{N-1})\, dR_{N-1} \end{aligned}\right\}. \quad (24)$$

In Rice's work, it is seen that the conditions listed in (24) are satisfied.

Remarks

It has been shown that if a Gaussian noise has a power spectrum such as obtained from a single-tuned narrow-band RLC filter with white noise at the input, the envelope constitutes a Markoff process. It should be mentioned that not only the envelope, but all the one-to-one zero-memory functions of this envelope will also be Markoff. In particular, if one defines the instantaneous power of the noise as one half the square of the instantaneous envelope, this is also Markoff. We have thus arrived at a new family of Markoff processes. Although no simpler proof is apparent, it is not suprising that the envelope of the noise with the given power spectrum should be Markoff. For the correlation matrix that results from this spectrum, the inphase and quadrature components of the narrow-band noise are independent Gaussian variates, and by Doob's[5] result these components themselves constitute identical Markoff processes. The envelope formed from the squares of these components might therefore be expected to be Markoff.

Acknowledgment

The author is indebted to Dr. Robert Price of M.I.T. Lincoln Laboratory for many stimulating discussions, helpful advice and suggestions, and for reading the manuscript of this paper.

[5] J. L. Doob, "The Brownian movement and stochastic equations," *Ann. Math.*, vol. 43, pp. 351–369; April, 1942. (As quoted by Wang and Uhlenbeck, *op. cit.*)

Prediction and Filtering for Random Parameter Systems*

F. J. BEUTLER†

Summary—**This work generalizes the Wiener-Kolmogorov theory of optimum linear filtering and prediction of stationary random inputs. It is assumed here that signal and noise have passed through a random device before being available for filtering and prediction. A random device is a unit whose behavior depends on an unknown parameter for which an *a priori* probability distribution is given.**

A number of engineering applications are cited. Two of these are worked out in some detail to illustrate the optimization procedure.

I. Introduction

THE transfer function describing a linear time invariant system is often dependent on its environment or its application. Since utilization of equipment is not always precisely known at the time of its design, the designer must take into account the ensemble of situations with which the system might be faced. The inevitable result is a compromise design, perhaps intuitively slanted toward those situations which the system will encounter most frequently.

This paper presents a systematic technique of mean square optimum design for systems containing random components whose ensemble characteristics are known. It is assumed that the system to be optimized is presented with a (wide sense) stationary input which has been passed through a component with transfer function $H(\omega, \gamma)$; γ is a random variable for which the distribution function is given.

Before beginning the analysis, several examples are given where such a method is useful. As a first example, consider an amplitude-modulated ground-to-ground

* Manuscript received by the PGIT, June 4, 1958. This work was conducted in part by Project MICHIGAN under Dept. of the Army Contract DA-36-039 SC-52654, administered by the U. S. Army Signal Corps.

† Dept. of Aero. Eng., University of Michigan, Ann Arbor, Mich.

communications link. Here the ground reflections depend on the transmission distance, nature of terrain, height of antennas above ground, and the modulating frequency. The problem is to design a filter which removes the reflection components of the incoming signal, as well as receiver noise and ground clutter. Evidently, the designer does not know in advance the spectrum of the ground reflections or the signal power at the receiver, especially when the communications link is intended to be versatile. He can determine, however, the probability of its use in each possible mode and type of terrain.

A second example is concerned with an antiaircraft gun in which the chief problem is one of predicting the position of the target over the time required for the shell to reach the proper altitude. This problem has been treated by Wiener. However, it may be assumed that random variations in air density and muzzle velocity of the shell affect the time of flight. Then the true time of flight (and thus the prediction time) is unknown and can only be determined statistically.

As a third application, suppose that a quantity is measured by a transducer located in an inaccessible or dangerous place. Although the transducer characteristics change with time, recalibration is impossible. Nevertheless, the ensemble behavior of such transducers can be established through experimentation, and this information used to treat the transducer data optimally by the methods of this paper.

II. Derivation of the Optimum Filter

The filter problem under consideration perhaps is best summarized by the flow diagram in Fig. 1. Assume for the moment that the noise $n(t) = 0$. The signal $x(t)$ is taken to be wide sense stationary and continuous in the mean.[1] The desired or ideal output, $y(t)$, is obtained by subjecting $x(t)$ to an ideal operator $u(\omega)$, defined in the frequency (but not necessarily the time) domain. The operator $k(\omega)$ is the optimum filter which is sought. It is restricted to being the transform of a physically realizable weighting function. This means that the ouput of the filter $k(\omega)$ cannot anticipate the input or, what is equivalent, the weighting function to which $k(\omega)$ is related must be zero for negative argument.

A novel feature of the filter problem depicted by Fig. 1 is that the filter cannot act on the signal itself, but on a randomly distorted version of the signal. This is the effect of the random parameter system $H(\omega, \gamma)$ through which $x(t)$ is passed before being available for filtering. An alternative formulation, of course, is that the optimum filter is always cascaded with $H(\omega, \gamma)$; that is, part of the filter itself is fixed in a random manner.

Since the input $z(t, \gamma)$ to the optimum filter has a spectrum which is dependent on γ, the z process is (in general) not ergodic, so that the entire optimization procedure must be expressed in terms of ensembles. This approach also leads to more general results; for instance, the ideal operator $u(\omega)$ need not have a time domain interpretation. Dealing with ensemble (rather than time) averages also makes it easy to prove the existence and uniqueness of the optimum filter and to show that the minimum mean square error can always be approached as closely as desired by stable lumped parameter networks.[2]

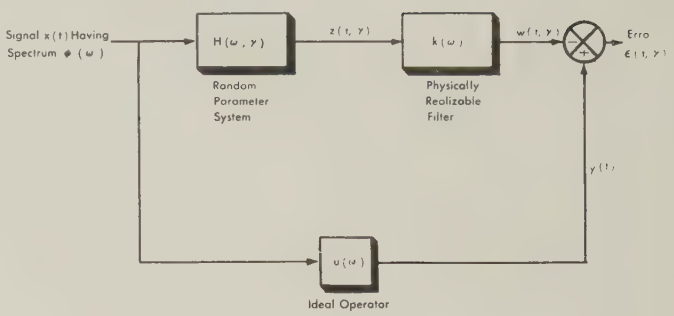


Fig. 1—Simplified random parameter filter problem.

A signal $x(t)$ can be represented by

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t}\, dX(\omega) \tag{1}$$

if $x(t)$ is a wide sense stationary process which is continuous in the mean.[3] Here $X(\omega)$ is a process with orthogonal increments for which

$$E \mid X(\omega_2) - X(\omega_1) \mid^2 = 2\pi \int_{\omega_1}^{\omega_2} \Phi(\omega)\, d\omega \tag{2}$$

where $\Phi(\omega)$, the spectral density of $x(t)$, is assumed to exist.

Now $y(t)$ is the result of operating on $x(t)$ with ideal operator $u(\omega)$. The representation for $y(t)$ is then

$$y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} u(\omega)\, dX(\omega) \tag{3}$$

in which any $u(\omega)$ which satisfies

$$\int_{-\infty}^{\infty} \Phi(\omega) \mid u(\omega) \mid^2 d\omega < \infty \tag{4}$$

is admissible.[4] The class of admissible $u(\omega)$ includes prediction [*i.e.*, $u(\omega) = e^{i\omega a}$ with $a > 0$], differentiation [if $\int_{-\infty}^{\infty} \omega^2\, \Phi(\omega)\, d\omega < \infty$], as well as any $u(\omega)$ which is the transform of an absolutely integrable weighting function $U(t)$. In the latter case,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} u(\omega)\, dX(\omega) = \int_{-\infty}^{\infty} x(\tau) U(t - \tau)\, d\tau \tag{5}$$

with probability 1; this establishes the mathematical link between time and frequency domains.

The signal $z(t, \gamma)$ into the optimum filter is expressed as

[1] See J. L. Doob, "Stochastic Processes," John Wiley and Sons, Inc., New York, N. Y., pp. 95, 518; 1953. These are minimum restrictions for optimum filtering.
N. Wiener requires stronger conditions in "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, N. Y.; 1949.

[2] F. J. Beutler, "A generalization of Wiener optimum filtering and prediction," Doctoral dissertation, California Inst. Tech., Pasadena, Calif., pp. 19–20; 1957.

[3] Doob, *op. cit.*, pp. 527–528.

[4] That is, $u(\omega)$ must be such that $y(t)$ has a finite mean square.

$$z(t, \gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} H(\omega, \gamma)\, dX(\omega) \tag{6}$$

which is analogous to (3). The optimum filter output $w(t, \gamma)$ represents a linear operation $k(\omega)$ on $z(t, \gamma)$ so that

$$w(t, \gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} H(\omega, \gamma) k(\omega)\, dX(\omega). \tag{7}$$

Since the error $\epsilon(t, \gamma)$ is defined as $y(t) - w(t, \gamma)$, we have from (3) and (7):

$$\epsilon(t, \gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} [u(\omega) - H(\omega, \gamma) k(\omega)]\, dX(\omega). \tag{8}$$

The mean-square-error ϵ^2 is determined by averaging $|\epsilon(t, \gamma)|^2$ over the ensemble of signals $x(t)$, and averaging also over γ. These steps may be performed in succession. Thus, (8) is used to average respective to $x(t)$, and to obtain the result

$$\epsilon^2 = \underset{\gamma}{E}\{\underset{x}{E} \mid \epsilon(t, \gamma) \mid^2\}$$

$$= \underset{\gamma}{E}\left\{\int_{-\infty}^{\infty} \mid u(\omega) - H(\omega, \gamma) k(\omega) \mid^2 \Phi(\omega)\, d\omega\right\}. \tag{9}$$

For convenience in notation, expectations on γ are written as

$$\underset{\gamma}{E}[H(\omega, \gamma)] = \int_{-\infty}^{\infty} H(\omega, \gamma)\, dF(\gamma) = H(\omega) \tag{10}$$

and

$$\underset{\gamma}{E} \mid H(\omega, \gamma) \mid^2 = \int_{-\infty}^{\infty} \mid H(\omega, \gamma) \mid^2 dF(\gamma) = G(\omega) \tag{11}$$

in which $F(\gamma)$ is the distribution function for the random variable γ. Expansion of (9) followed by a regrouping of terms then yields

$$\epsilon^2 = \int_{-\infty}^{\infty} \left| k(\omega) - \frac{u(\omega)\overline{H(\omega)}}{G(\omega)} \right|^2 G(\omega)\Phi(\omega)\, d\omega$$

$$+ \int_{-\infty}^{\infty} \left[1 - \frac{\mid H(\omega) \mid^2}{G(\omega)}\right] \mid u(\omega) \mid^2 \Phi(\omega)\, d\omega. \tag{12}$$

The second integral in (12) is seen to be independent of any optimization of $k(\omega)$. This term is always nonnegative by the Schwarz inequality; in fact, $|H(\omega)|^2 = G(\omega)$ if and only if there is an $\tilde{H}(\omega)$ such that $\tilde{H}(\omega) = H(\omega, \gamma)$ independently of γ. In physical terms, a random transfer function element cascaded with the filter $k(\omega)$ always leaves a residual mean square error. Because the integral does not depend on $k(\omega)$, the residual error remains even if $k(\omega)$ is not restricted to realizable filters.

Consider now the first integral of (12). The optimization problem is solved if we find that $k(\omega)$ which minimizes the integral. The method of solution proposed here places the minimization of the integral within the context of Wiener filtering and prediction. Suppose that a process $r(t)$ has spectral density $\Phi(\omega)G(\omega)$, and that the desired output of the system is obtained by subjecting $r(t)$ to the transfer function $u(\omega)\overline{H(\omega)}/G(\omega)$. The actual output is generated

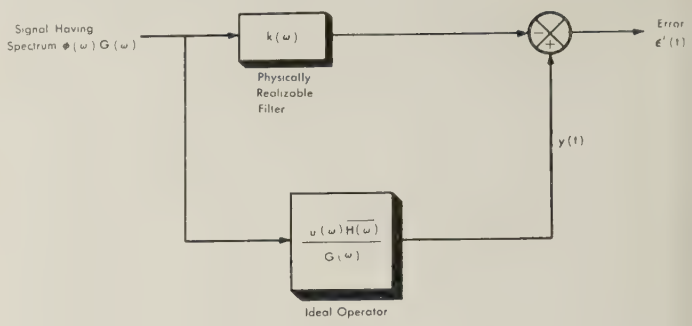


Fig. 2—Wiener problem having the same optimum filter as in Fig. 1.

by passing $r(t)$ through a physically realizable filter with transfer function $k(\omega)$. These relationships are depicted in Fig. 2. The mean square error for this system is

$$\epsilon'^2 = \int_{-\infty}^{\infty} \left| k(\omega) - \frac{u(\omega)\overline{H(\omega)}}{G(\omega)} \right|^2 \Phi(\omega)G(\omega)\, d\omega \tag{13}$$

which corresponds precisely to the first integral in (12). Thus the problems are identical and the Wiener technique is applicable, if only $\Phi(\omega)G(\omega)$ is factorable.

The question of the factorization of $\Phi(\omega)G(\omega)$ is deferred for the moment; we assume that $\Phi(\omega) = \psi_1(\omega)\overline{\psi_1(\omega)}$ and $G(\omega) = G_1(\omega)\overline{G_1(\omega)}$. Here ψ_1 and G_1 are Fourier transforms of L_2 functions[5] which are zero for negative argument. The optimum Wiener filter for the problem of Fig. 2 (and thus for the problem of Fig. 1) then is known to be

$$k(\omega) = \frac{1}{2\pi\psi_1(\omega)G_1(\omega)}$$

$$\cdot \int_0^{\infty} e^{i\omega t}\left[\int_{-\infty}^{\infty} e^{i t\rho} \frac{\psi_1(\rho)\overline{H(\rho)}u(\rho)}{\overline{G_1(\rho)}}\, d\rho\right] dt. \tag{14}$$

The $k(\omega)$ determined by (14) is merely the projection of $u(\omega)\overline{H(\omega)}/G(\omega)$ on the space of realizable function, as shown by Doob.[6] Corresponding to $k(\omega)$ the mean square error ϵ'^2 is

$$\epsilon'^2 = \frac{1}{2\pi} \int_{-\infty}^{0} \left| \int_{-\infty}^{\infty} e^{i t\rho} \frac{\psi_1(\rho)\overline{H(\rho)}u(\rho)}{\overline{G_1(\rho)}}\, d\rho \right|^2 dt. \tag{15}$$

The latter computation follows directly from (13) and (14). It follows also that the total mean-square-error ϵ^2 for the original problem is determined as

$$\epsilon^2 = \frac{1}{2\pi} \int_{-\infty}^{0} \left| \int_{-\infty}^{\infty} e^{i t\rho} \frac{\psi_1(\rho)\overline{H(\rho)}u(\rho)}{\overline{G_1(\rho)}}\, d\rho \right|^2 dt$$

$$+ \int_{-\infty}^{\infty} \Phi(\omega) \mid u(\omega) \mid^2 \left[1 - \frac{\mid H(\omega) \mid^2}{G(\omega)}\right] d\omega \tag{16}$$

from (12) and (15).

We return now to the factorization problem. In the first place, both Φ and G must be factorable if their product is to be factored. According to a theorem of

[5] A function $f(t)$ is said to be in L_p if $\int_{-\infty}^{\infty} |f(t)|^p\, dt < \infty$.

[6] This is a rephrasing of the filtering and prediction problem in terms of Hilbert spaces. Such a treatment permits more general results in a rigorous fashion. See Doob, *op. cit.*, ch. 12.

Paley and Wiener,[7] a nonnegative function in L_1 [for example, $\Phi(\omega)$] is factorable if and only if

$$\int_{-\infty}^{\infty} \frac{\log \Phi(\omega)}{1+\omega^2}\, d\omega > -\infty. \tag{17}$$

The condition (18) is also necessary and sufficient to insure that the entire future of $x(t)$ is not precisely predictable from its past.[8] Therefore, it can be assumed that (17) is true because it makes no sense to predict or filter deterministic functions.

As for $G(\omega)$, assume that $H(\omega, \gamma)$ is the Fourier transform of $h(t, \gamma)$ in L_1, where $h(t, \gamma) = 0$ when $t < 0$ for all γ. Then $H(\omega)$ is the transform of a function which is zero for negative t, and therefore

$$\int_{-\infty}^{\infty} \frac{\log \mid H(\omega) \mid^2}{1+\omega^2}\, d\omega > -\infty \tag{18}$$

by the above-mentioned theorem of Paley and Wiener. But by the Schwarz inequality, $G(\omega) \geq \mid H(\omega) \mid^2$, so that (18) implies

$$\int_{-\infty}^{\infty} \frac{\log G(\omega)}{1+\omega^2}\, d\omega > -\infty. \tag{19}$$

This is exactly the condition (17) applied to $G(\omega)$. Then $G(\omega)$ is factorable.

It should be noted that $G(\omega)$ is not rational even if $H(\omega, \gamma)$ is rational, unless γ can assume only a finite number of values. This makes the factorization more difficult, but factors can always be constructed[9] if $h(t, \gamma)$ meets the requirements of the preceding paragraph.

III. Filtering and Prediction of a Signal with Noise

Fig. 3 illustrates an extension of the preceding analysis. A noise input has been added, and the noise passed through a random system having a transform $L(\omega, \mu)$. The signal input is passed through $H(\omega, \gamma)$, as before.

If the spectral density of the signal and noise are denoted by $\Phi_s(\omega)$ and $\Phi_n(\omega)$, respectively, the spectral density of the total input is

$$\Phi(\omega) = \Phi_s(\omega)G(\omega) + \Phi_n(\omega)M(\omega). \tag{20}$$

Here $M(\omega)$ is defined by $M(\omega) = \overset{E}{\mu} \mid L(\omega, \mu) \mid^2$. γ and μ need not be independent; (20) only assumes that signal and noise are uncorrelated.

If $M(\omega)$ meets the same conditions as those previously imputed to $G(\omega)$, it is possible to factor $\Phi(\omega)$ whenever signal and/or noise spectra satisfy (17). Suppose, then,

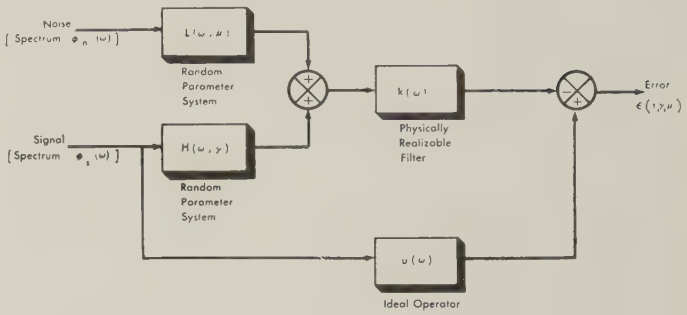


Fig. 3—Generalized random parameter filter problem.

that $\Phi(\omega) = \chi_1(\omega)\overline{\chi_1(\omega)}$. In that case, the optimum filter is found to be

$$k(\omega) = \frac{1}{2\pi\chi_1(\omega)} \int_0^{\infty} e^{-i\omega t} \left[\int_{-\infty}^{\infty} e^{it\rho}\, \frac{\Phi_s(\rho)\overline{H(\rho)}u(\rho)}{\overline{\chi_1(\rho)}}\, d\rho \right] dt \tag{21}$$

by methods identical with those of the preceding section. Likewise, a mean-square-error computation yields.

$$\epsilon^2 = \frac{1}{2\pi} \int_{-\infty}^{0} \left| \int_{-\infty}^{\infty} e^{it\omega}\, \frac{\Phi_s(\omega)\overline{H(\omega)}u(\omega)}{\overline{\chi_1(\omega)}}\, d\rho \right|^2 dt$$

$$+ \int_{-\infty}^{\infty} \Phi_s(\omega) \mid u(\omega) \mid^2 \left[1 - \frac{\Phi_s(\omega) \mid H(\omega) \mid^2}{\Phi(\omega)} \right] d\omega. \tag{22}$$

As might be expected, a reduction to ordinary Wiener filtering is obtained if $H(\omega, \gamma)$ is not truly random. For instance, $H(\omega, \gamma)$ may be independent of γ, or γ may assume a specified value with probability 1. It follows that $G(\omega) = \mid H(\omega) \mid^2$, since now averaging over γ and squaring $H(\omega, \gamma)$ may be freely interchanged. The optimization problem in question then is exactly equivalent to the following Wiener problem: let the filter input consist of a signal having spectral density $\Phi_s(\omega) \mid H(\omega) \mid^2$ and noise with spectral density $\Phi_n(\omega)M(\omega)$, and make the desired (ideal) operator on the signal equal to $u(\omega)/H(\omega)$. Indeed, the Wiener filter for the problem is given by (21), and the least mean square error by (22), providing that the substitution $G(\omega) = \mid H(\omega) \mid^2$ is made.

The relationship between the filtering proposed here and Wiener filtering becomes even more apparent when no noise is present. This is the case treated in the preceding section. There, part of the numerator and denominator in the integrand of (15) are seen to cancel. In the mean-square-error expression (16), the second integral disappears entirely, since now $\mid H(\omega) \mid^2/G(\omega) = 1$. In other words, a filter is designed for a signal spectrum for which the ideal operator is to be $u(\omega)/H(\omega)$.

A case of special interest is that of pure prediction or lag filters in the presence of noise. This means that $u(\omega) = e^{i\omega a}$; a positive implies prediction, while a negative calls for a lagging filter. We now have

$$k(\omega) = \frac{1}{2\pi\chi_1(\omega)} \int_0^{\infty} e^{-i\omega t} \left[\int_{-\infty}^{\infty} e^{i(t+a)\rho}\, \frac{\Phi_s(\rho)\overline{H(\rho)}}{\overline{\chi_1(\rho)}}\, d\rho \right] dt \tag{23}$$

for the optimum filter. A simple computation yields

[7] R. E. A. C. Paley and N. Wiener, "Fourier Transforms in the Complex Domain," American Mathematical Society Colloqium Publication, vol. 19, pp. 17–20; 1934.

[8] A. N. Kolmogorov, "Interpolation and extrapolation von stationären zufalligen Folgen," *Bull. Acad. Sci. U.S.S.R., Ser. Math.*, vol. 5, pp. 5–14; 1941.

[9] A method of accomplishing this construction is given by N. Levinson, "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," *J. Math. Phys.*, vol. 26, pp. 110–119; July, 1947.

$$\epsilon^2 = \frac{1}{2\pi}\int_{-\infty}^{0}\left|\int_{-\infty}^{\infty} e^{i(t+a)\omega}\frac{\Phi_s(\omega)\overline{H(\omega)}}{\overline{\chi_1(\omega)}}\,d\omega\right|^2 dt + \int_{-\infty}^{\infty}\Phi_s(\omega)\left[1-\frac{\Phi_s(\omega)\mid H(\omega)\mid^2}{\Phi(\omega)}\right]d\omega \qquad (24)$$

as the mean square error. The mean square error is thus a monotone nondecreasing function of a. If the prediction interval is infinite, the optimum filter has zero for its transfer function, and the mean square error is given by the input signal power. When an infinite lag is permitted,[10] we have $k(\omega) = \Phi_s(\omega)\overline{H(\omega)}/\overline{\chi_1(\omega)}$, and the only error is due to the randomness of the system, *viz.*,

$$\epsilon^2 = \int_{-\infty}^{\infty}\Phi_s(\omega)\left[1-\frac{\Phi(\omega)\mid H(\omega)\mid^2}{\Phi(\omega)}\right]d\omega. \qquad (25)$$

IV. Examples

The first example studied now is concerned with a communication link in which it is desired that the transmitted message $x(t)$ be exactly reproduced at the receiver in real time. This means that a predicting filter is necessary at the receiver to obtain the best estimate of $x(t)$ from the present and past of $x(t-\gamma)$; γ is the time delay between transmission and reception of the message. Such a delay is the result of propagation time (particularly important in acoustic devices), and perhaps lags occurring in the modulation and/or demodulation process.

The future utilization of a mobile unit is not generally known at the time of its design, so that the designer may well wish to regard the delay γ as a random variable whose probability distribution he can determine from empirical data gathered on similar apparatus. Furthermore, tolerances and aging of electronic components will also affect γ in a random manner.

It is seen from the preceding discussion that the optimum filter must treat the signal $x(t)$ after it has been subjected to the random delay operator $H(\omega,\gamma) = e^{-i\omega\gamma}$. Since noise is assumed absent, the results of Section II are applicable. In terms of the notation of that section, $u(\omega) = 1$ because the desired filter output is identical with the original signal.

If the probability distribution function of γ is given by $F(\gamma)$, $H(\omega)$ may be evaluated as

$$H(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\gamma}\,dF(\gamma) = \overline{\phi(\omega)} \qquad (26)$$

where $\phi(\omega)$ is, by definition, the characteristic function[11] for the random variable γ. That $G(\omega) = 1$ is seen from the fact that $\mid H(\omega,\gamma)\mid$ is unity for every ω and γ.

Substitution in (14) yields the result

$$k(\omega) = \frac{1}{2\pi\psi_1(\omega)}\int_0^{\infty} e^{-i\omega t}\left[\int_{-\infty}^{\infty} e^{it\rho}\psi_1(\rho)\phi(\rho)\,d\rho\right]dt \qquad (27)$$

for the optimum filter. The mean square error for this filter is

$$\epsilon^2 = \frac{1}{2\pi}\int_{-\infty}^{0}\left|\int_{-\infty}^{\infty} e^{it\omega}\psi_1(\omega)\phi(\omega)\,d\omega\right|^2 dt + \int_{-\infty}^{\infty}\Phi(\omega)[1-\mid\phi(\omega)\mid^2]\,d\omega \qquad (28)$$

from (16).

An interesting variation of the above obtains the solution to the antiaircraft gun problem described in the Introduction. If the shell is fired at time t, we must be able to predict the target position at time $t+\gamma$, *i.e.*, the time at which the shell reaches the altitude of the target. In other words, a prediction filter is required to compensate as nearly as possible for the delay γ inherent in the system.

If it were possible to designate γ in advance, the filter should be designed by Wiener's method; however, the time of flight to the altitude of the target aircraft may vary from occasion to occasion, so that γ is known only in a statistical sense.

For the purposes of this illustration, a one-dimensional solution of the problem will be offered. The target position is denoted by x, while v is taken to represent its velocity. The hell is fired at time t, and $x(\tau)$ is known for all $\tau \le t$. It is desired to make a least square estimate of $x(t+\gamma)$, γ being a random variable with some probability distribution function $F(\gamma)$.

It has sometimes been assumed that v is stationary with spectral density[12]

$$\Phi(\omega) = \frac{1}{1+\omega^2}. \qquad (29)$$

The form (29) raises the difficulty that x (the integral of v) is nonstationary, whereas our theory applies only to stationary inputs. The stationarity problem is surmounted as follows. Since $x(t)$ is already known when the shell is fired, only a prediction of $x(t+\gamma) - x(t)$ need be made. This difference is produced by subjecting the stationary process v to the operator $e^{i\omega\gamma}-1/i\omega$. To prove that this operator on v gives the desired quantity, consider the representation for $v(t)$:

$$v(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega t}\,dV(\omega). \qquad (30)$$

Applying the operator in question then yields

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega t}\left(\frac{e^{i\omega\gamma}-1}{i\omega}\right)dV(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\omega t}\left[\int_0^{\gamma} e^{i\omega\tau}\,d\tau\right]dV(\omega) = \int_0^{\gamma} v(t+\tau)\,d\tau = x(t+\gamma) - x(t). \qquad (31)$$

[10] An infinite lag becomes applicable whenever recorded data are to be reduced at a later date.

[11] A definition and some of the properties of characteristic functions are given by Doob, *op. cit.*, pp. 37 ff.

[12] This form of the spectral density is justified in H. M. James, N. B. Nichols, and R. S. Phillips, "Theory of Servomechanisms," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., no. 25, pp. 300–304; 1947.

The antiaircraft problem now can be defined in terms of Fig. 1. The system input is now $v(t)$, and $H(\omega, \gamma) = 1$. Indeed, it is the ideal operator which now varies according to target altitude. In recognition of this fact, this operator now is called $u(\omega, \gamma)$. It follows from (31) that

$$u(\omega, \gamma) = \frac{e^{i\omega\gamma} - 1}{i\omega}. \tag{32}$$

Because the input is velocity rather than position, the optimum filter $k(\omega)$ must be thought of as having a velocity input. Should the input actually consist of the position, the filter need merely be multiplied by $i\omega$ to retain its optimum properties.

The filtering error is again the difference between ideal and actual output, that is,

$$\epsilon(t, \gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} \left[\frac{e^{i\omega\gamma} - 1}{i\omega} - k(\omega) \right] dV(\omega). \tag{33}$$

Averaging the squared error over γ as well as v gives the result

$$\epsilon^2 = \int_{-\infty}^{\infty} dF(\gamma) \int_{-\infty}^{\infty} \Phi(\omega) \left| \frac{e^{i\omega\gamma} - 1}{i\omega} - k(\omega) \right|^2 d\omega. \tag{34}$$

The procedure of minimizing over the class of realizable function is the same as before. We obtain

$$k(\omega) = \frac{1}{2\pi\psi_1(\omega)} \int_0^{\infty} e^{-i\omega t} \left[\int_{-\infty}^{\infty} e^{i t \rho} \psi_1(\rho) \frac{1 - \phi(\rho)}{i\rho} d\rho \right] dt \tag{35}$$

where $\phi(\omega)$ is again the characteristic function of γ. We note that (35) can be simplified considerably by resolving the inner integral through contour integration in the upper half plane.[13] If γ has moments higher than the first [say $\int_0^{\infty} \gamma^{1+\delta}\, dF(\gamma)$ exists for come $\delta > 0$] there is no pole at the origin, so that only the residue at $z = i$ is considered.[14] The final result is

$$k(\omega) = 1 - \phi(i) = 1 - \int_0^{\infty} e^{-\gamma}\, dF(\gamma). \tag{36}$$

This result is entirely consistent with a Wiener prediction filter; should $\gamma = \alpha$ with probability 1, $k(\omega) = 1 - e^{-\alpha}$ from (36).

The mean square error which corresponds to the optimum filter (36) is easily computable through use of (34) and (36). This error is

$$\epsilon^2 = \pi[1 - \phi(i)]^2 + \int_{-\infty}^{\infty} \frac{1}{\omega^2 + 1} [1 - |\phi(\omega)|^2]\, d\omega \tag{37}$$

which also reduces to the Wiener result if γ is certain rather than random.

For a comparison between the optimum and any other filter $\hat{k}(\omega)$, the difference in mean square errors is expressed by

$$\Delta\epsilon^2 = \int_{-\infty}^{\infty} \Phi(\omega) \, | \, k(\omega) - \hat{k}(\omega) \, |^2 \, d\omega. \tag{38}$$

In particular, our filter outperforms the fixed prediction interval (Wiener) filter by[15]

$$\Delta\epsilon^2 = \frac{1}{2\pi} \int_0^{\infty} \left| \int_{-\infty}^{\infty} e^{i t \omega} \frac{\psi_1(\omega)[\phi(\omega) - e^{i a \omega}]}{i\omega} d\omega \right|^2 dt \tag{39}$$

where the Wiener filter has been designed on the assumption that $\gamma = a$. In the event that $\phi(\omega) = 1/1 + \omega^2$, (39) specializes to

$$\Delta\epsilon^2 = \pi \left| e^{-a} - \int_0^{\infty} e^{-\gamma}\, dF(\gamma) \right|^2 \tag{40}$$

which may be obtained by contour integration. Thus the Wiener filter is as effective (in the mean square sense) as our filter if and only if $\gamma = a$ with probability 1.

V. Acknowledgment

The author expresses his appreciation to Prof. C. R. DePrima of the California Institute of Technology. Many of the results of this paper are the consequence of his invaluable suggestions.

[13] Since γ must be greater than zero, $\phi(z)$ is regular in the upper half plane.

[14] The existence of these moments is a sufficient condition for the existence of the integral (34), for then we have the expansion $\phi(\omega) = 1 + i\omega m + 0(\omega^{1+\delta})$ in which m is the mean of γ. Note, however, that the integral converges when γ has the Couchy distribution for which not even the first moment exists.

[15] Eq. (39) is obtained from (38) by substituting the required computation for k and $\hat{k}$, and then applying Parseval's relation.

Correspondence

A Criterion for the Diagonal Expansion of a Second-Order Probability Distribution in Orthogonal Polynomials*

In a recent paper,[1] Barrett and Lampard introduced an expansion for second-order probability distributions which expresses such a distribution as a double series involving orthogonal polynomials associated with the corresponding first-order probability distributions. Several interesting consequences were derived for the class, Λ, consisting of all second-order distributions having a *diagonal* expansion of the form

$$p(x_1; x_2) = p_1(x_1)p_2(x_2) \cdot \sum_{n=0}^{\infty} a_n \theta_n^{(1)}(x_1)\theta_n^{(2)}(x_2), \qquad (1)$$

but it was stated that the authors had not found what general restrictions must be placed on $p(x_1; x_2)$ in order that it may belong to Λ.

If the conditional moments,

$$\int \frac{p(x_1; x_2)}{p_1(x_1)} x_2^k \, dx_2$$

and

$$\int \frac{p(x_1; x_2)}{p_2(x_2)} x_1^k \, dx_1$$

are denoted by $m_k^{(1)}(x_1)$ and $m_k^{(2)}(x_2)$, respectively ($k = 0, 1, 2, \ldots$), then a characterization of the class A is given by the following theorem.

Theorem: Assume that $p(x_1; x_2)$ can be represented as a double series in the associated polynomials. Then $p(x_1; x_2)$ belongs to Λ if, and only if, the quantities $m_k^{(1)}(x_1)$ and $m_k^{(2)}(x_2)$ are polynomials (in their respective variables) of degree less than or equal to k, for each positive integral value of k.

Proof: If $p(x_1; x_2)$ has the form (1), then

$$m_k^{(1)}(x_1) = \sum_{n=0}^{\infty} a_n \theta_n^{(1)}(x_1) \cdot \int \theta_n^{(2)}(x_2) x_2^k p_2(x_2) \, dx_2$$

$$= \sum_{n=0}^{k} a_n \theta_n^{(1)}(x_1) \cdot \int \theta_n^{(2)}(x_2) x_2^k p_2(x_2) \, dx_2$$

since any orthogonal polynomial is orthogonal to *every* polynomial of lower degree on the same interval and with respect to the same weighting function.[2] The same argument shows that $m_k^{(2)}(x_2)$ is also a polynomial of maximum degree k.

Conversely, assume $m_k^{(1)}(x_1)$ and $m_k^{(2)}(x_2)$ are each polynomials of degree less than or equal to k for $k = 0, 1, 2, \cdots$. Then for fixed n,

$$\pi_n^{(1)}(x_1) \equiv \int \frac{p(x_1; x_2)}{p_1(x_1)} \theta_n^{(2)}(x_2) \, dx_2$$

is a polynomial of degree $\leq n$, since it is a linear combination of $m_0^{(1)}(x_1), \cdots, m_n^{(1)}(x_1)$. Similarly,

$$\pi_m^{(2)}(x_2) \equiv \int \frac{p(x_1; x_2)}{p_2(x_2)} \theta_m^{(1)}(x_1) \, dx_1$$

is a polynomial of degree $\leq m$.

In the expansion

$$p(x_1; x_2) = p_1(x_1)p_2(x_2) \cdot \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} \theta_m^{(1)}(x_1)\theta_n^{(2)}(x_2) \qquad (2)$$

the coefficients a_{mn} are given by

$$a_{mn} = \iint p(x_1; x_2) \theta_m^{(1)}(x_1) \theta_n^{(2)}(x_2) \, dx_1 \, dx_2$$

$$= \int p_1(x_1) \pi_n^{(1)}(x_1) \theta_m^{(1)}(x_1) \, dx_1$$

$$= 0 \quad \text{for} \quad m > n, \qquad (3)$$

using the previously stated property of orthogonal polynomials. Also,

$$a_{mn} = \int p_2(x_2) \pi_m^{(2)}(x_2) \theta_n^{(2)}(x_2) \, dx_2$$

$$= 0 \quad \text{for} \quad m < n,$$

Thus, for arbitrary m and n, $a_{mn} = 0$ for $m \neq n$; that is, the expansion is diagonal. *Q.E.D.*

Example: To illustrate the use of this criterion, consider the familiar Gaussian process, where

$$p(x_1; x_2) = \frac{1}{2\pi\sigma^2\sqrt{1 - \rho^2}} \cdot \exp\left[-\frac{1}{2} \left\{ \frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\sigma^2(1 - \rho^2)} \right\} \right]$$

and

$$p_1(x_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{x_1^2}{\sigma^2} \right].$$

Then,

$$m_k^{(1)}(x_1) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi(1 - \rho^2)}} \cdot \exp\left[-\frac{1}{2} \left\{ \frac{(x_2 - \rho x_1)^2}{\sigma^2(1 - \rho^2)} \right\} \right] x_2^k \, dx_2$$

$$= \frac{1}{\sigma\sqrt{2\pi(1 - \rho^2)}} \cdot \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \left\{ \frac{\xi^2}{\sigma^2(1 - \rho^2)} \right\} \right] \cdot (\rho x_1 + \xi)^k \, d\xi$$

which, by inspection, is seen to be a polynomial in x_1 of maximum degree k. By symmetry, $m_k^{(2)}(x_2)$ is also a polynomial of maximum degree k and by the theorem, $p(x_1; x_2)$ has a diagonal expansion in orthogonal polynomials. Note that the relevant polynomials do not have to be known in order to apply the criterion. In this case, for instance, the polynomials are the Hermite polynomials and carrying out the actual expansion is equivalent to a derivation of Mehler's formula.[1] Since the results cited by Barrett and Lampard depend only on the diagonal property of the expansion and not on the explicit form of the associated polynomials, the above criterion is expedient in that it is phrased entirely in terms of the given distribution and does not require construction of the polynomials by the rather tedious Schmidt process.

It may also be of interest to note that the orthogonal polynomials used in the Barrett-Lampard expansion[1,3] are essentially equivalent to the moments of the two first-order distributions; that is, the moments may be derived from knowledge of the polynomials and conversely. The proof, which is straightforward, is left to the reader.

Lastly, it seems natural to ask whether $p(x_1; x_2)$ can have a *diagonal* expansion in terms of orthogonal polynomials different from the polynomials of the Barrett-Lampard expansion. If the zeroth degree polynomials, $\phi_0^{(1)}(x_1)$ and $\phi_0^{(2)}(x_2)$, of an alternative polynomial representation are each taken as unity, then the polynomials will be those of the Barrett-Lampard expansion. The proof, which presents no difficulties, shows that the Barrett-Lampard expansion is unique in the sense that any diagonal expansion of $p(x_1; x_2)$ in orthogonal polynomials must be of this type.

John L. Brown, Jr.
Ordnance Res. Lab.
Pennsylvania State University
University Park, Pa.

* Received by the PGIT, January 30, 1958.

[1] J. F. Barrett and D. G. Lampard, "An expansion for some second-order probability distributions and its application to noise problems," IRE Trans. on Information Theory, vol. IT-1, pp. 10–15; March, 1955. (The notation of this reference is adopted in the present note.)

[2] D. Jackson, "Fourier Series and Orthogonal Polynomials," Math. Assoc. of America, p. 154; 1941.

[3] J. L. Brown, Jr., "On a cross-correlation property for stationary random processes," IRE Trans. on Information Theory, vol. IT-3, pp. 28–31; March, 1957. (General properties of the nondiagonal expansion are stated.)

On Signal Parameter Estimation*

The purpose of this note is to point out a simple but interesting result which is obtained when the ideas of Statistical Decision Theory[1] are applied to estimate the parameters of signals represented by a linear expansion of orthogonal functions.[2]

Suppose a signal $S(t)$ is represented by an expansion of the form

$$S(t) = \sum_{j=1}^{\infty} a_j \phi_j(t) \tag{1}$$

and vector $\mathbf{S} = (a_1, a_2, \cdots, a_j \cdots)$ where the ϕ_j are orthonormal on some interval which completely spans the interval (O, T) and $S(t)$ is assumed to exist only in (O,T). Let $S(t)$ be imbedded in additive white, Gaussian noise $N(t)$ whose spectral density is N_0 watts/rad/second. The combination is the data, $V(t)$:

$$V(t) = S(t) + N(t).$$

A Bayes estimator, $\gamma(\mathbf{V})$, of $\mathbf{S}$ on the basis of the data $V(t)$ is desired in the form of a set of estimates of its expansion coefficients. Two cost functions are considered:

a) $C(\mathbf{S}, \boldsymbol{\gamma}) = || \mathbf{S} - \boldsymbol{\gamma} ||^2$,

the quadratic cost function;

b) $C(\mathbf{S}, \boldsymbol{\gamma}) = A - \delta(\mathbf{S} - \boldsymbol{\gamma})$,

the "simple" cost function. (A is a positive constant and δ is the Dirac delta function.)

The quadratic cost function a) provides a cost which is equal to the squared error of the estimate. The simple cost function b) states that the cost of any incorrect estimate is A while that of a correct estimate is $-\infty$. These cost functions are discussed more fully in Middleton and Van Meter.[3]

The optimum estimator for the quadratic cost function is from (4.8) of Middleton and Van Meter.[1]

$$\gamma_\sigma^*(\mathbf{V}) = \frac{\int_\Omega \mathbf{S}\sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S})\, d\mathbf{S}}{\int_\Omega \sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S})\, d\mathbf{S}}. \tag{2}$$

$\sigma(\mathbf{S})$ is the *a priori* distribution of $\mathbf{S}$ in signal space Ω. Since $\mathbf{S}$ here is represented by (1), $\sigma(\mathbf{S})$ is written in terms of the probability densities of the individual parameters, a_j. These may or may not be independent of each other.

Let $V(t)$ also be represented by an expansion of the form

$$V(t) = \sum_{j-1}^{\infty} b_j \phi_j(t) \quad 0 \le t \le T \tag{3}$$

and vector $\mathbf{V} = (b_1, b_2, \cdots, b_j, \cdots)$.

For white, Gaussian noise

$$N(t) = \sum_{j=1}^{\infty} c_j \phi_j(t)$$

$$\mathbf{N} = (c_1, c_2, \cdots, c_j, \cdots)$$

where the c_j are normally distributed with mean, $\mu_j = 0$ and variance, $\sigma_j^2 = N_0$.

Eq. (2) is a vector relationship. For the kth component we have

$$[\gamma_\sigma^*(\mathbf{V})]_k = \frac{\int_\Omega a_k \sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S})\, da_k}{\int_\Omega \sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S})\, d\mathbf{S}}. \tag{4}$$

The covariance matrix, $\mathbf{K}_N$, of the noise has terms defined by

$$[\mathbf{K}_N]_{jk} = \overline{c_j c_k} = N_0 \delta_{jk} \tag{5a}$$

and

$$[\mathbf{K}_N^{-1}]_{jk} = \frac{\delta_{jk}}{N_0}. \tag{5b}$$

Then $F(\mathbf{V} \mid \mathbf{S})$, the likelihood function, can be written

$$F(\mathbf{V} \mid \mathbf{S}) = \prod_{j=1}^{\infty} \frac{1}{\sqrt{2\pi N_0}} \cdot \exp\left[\frac{-(b_j - a_j)^2}{2N_0}\right]. \tag{6}$$

We now assume that $\sigma(\mathbf{S})$ can be written in the form

$$\sigma(\mathbf{S}) = \sigma_1(a_1)\sigma_2(a_2) \cdots \sigma_j(a_j) \cdots \tag{7}$$

This assumption is quite cautious and perhaps unrealistic in that all signal parameters are assumed to be independent of each other—knowledge of one is of no help in obtaining information about the others. With actual signals there is apt to be considerable interdependence of the parameters. In this sense the estimators that follow below assume the worst possible *a priori* distribution for $\mathbf{S}$.[4]

$$[\gamma_\sigma^*(\mathbf{V})]_k = \frac{\frac{1}{\sqrt{2\pi N_0}} \int_{-\infty}^{\infty} a_k \sigma_k(a_k) \exp\left[-\frac{(b_k - a_k)^2}{2N_0}\right] da_k}{\frac{1}{\sqrt{2\pi N_0}} \int_{-\infty}^{\infty} \sigma_k(a_k) \exp\left[-\frac{(b_k - a_k)^2}{2N_0}\right] da_k}. \tag{8}$$

$\gamma_\sigma^*(\mathbf{V})$ is simply the conditional expectation of $\mathbf{S}$ given the data $\mathbf{V}$.

If we also assume that the a_j are uniformly distributed, $\sigma_k(a_k)$ is a constant and (4) becomes

$$[\gamma_\sigma^*(\mathbf{V})]_k = \int_{-\infty}^{\infty} a_k \frac{1}{\sqrt{2\pi N_0}} \cdot \exp\left[\frac{-1}{2}\frac{(b_k - a_k)^2}{N_0}\right] da_k \tag{9}$$

or

$$[\gamma_\sigma^*(\mathbf{V})]_k = b_k. \tag{10}$$

That is, for a signal whose orthogonal components are statistically independent and uniformly distributed, the quadratic cost function yields a Bayes estimator for the k^{th} component of $\mathbf{S}$

$$[\gamma_\sigma^*(\mathbf{V})]_k = \int_0^T V(t)\phi_k(t)\, dt. \tag{11}$$

The optimum estimator $\hat{\mathbf{S}}$ for the simple cost function is [from (4.4a) of Middleton and Van Meter[1]] the maximum likelihood estimate defined by

$$\sigma(\hat{\mathbf{S}})F(\mathbf{V} \mid \hat{\mathbf{S}}) \ge \sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S}). \tag{12}$$

Again, with white, Gaussian background noise and $\sigma(\mathbf{S})$ defined as in (7)

$$\sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S}) = \prod_{j=1}^{\infty} \frac{1}{\sqrt{2\pi N_0}} \cdot \exp\left[-\frac{1}{2}\frac{(b_j - a_j)^2}{N_0}\right] \tag{13}$$

from which the k^{th} component of $\hat{\mathbf{S}}$ is simply

$$[\hat{\mathbf{S}}]_k = b_k = \int_0^T V(t)\phi_k(t)\, dt. \tag{14}$$

It is also true that the b_k are Minimax estimators (Section 4.4 of Middleton and Van Meter[1]) of the a_k.

If the a_j are independent and normally distributed with mean μ_{a_j}, variance $\sigma_{a_j}^2$, then (9) becomes

$$[\gamma_\sigma^*(\mathbf{V})]_k = \frac{\int_{-\infty}^{\infty} a_k \frac{1}{2\pi\sigma_{a_k}\sqrt{N_0}} \exp\left[-\frac{1}{2}\frac{(a_k - b_k)^2}{N_0} - \frac{1}{2}\frac{(a_k - \mu_{a_k})^2}{\sigma_{a_k}^2}\right] da_k}{\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_{a_k}\sqrt{N_0}} \exp\left[-\frac{1}{2}\frac{(a_k - b_k)^2}{N_0} - \frac{1}{2}\frac{(a_k - \mu_{a_k})^2}{\sigma_{a_k}^2}\right] da_k} \tag{15}$$

* Received by the PGIT, June 4, 1958.

[1] D. Middleton and D. Van Meter, "Detection and extraction of signals in noise from the point of view of statistical decision theory." parts I and II, *J. Soc. Indus. Appl. Math.*, vol. 3, p. 192; December, 1955, and vol. 4, p. 86; June, 1956.

[2] W. H. Huggins, "Signal theory," IRE TRANS. ON CIRCUIT THEORY, vol. CT-3, pp. 210–216; December, 1956.

[3] Middleton and Van Meter, *op. cit.*, part II, pp. 88.

[4] See also Middleton and Van Meter, *op. cit.*, part II, p. 97.

or

$$[\gamma_{\sigma}^{*}(\mathbf{V})]_k = \frac{(b_k\sigma_{a_k}^2 + \mu_{a_k}N_0)}{(\sigma_{a_k}^2 + N_0)}. \tag{16}$$

Similarly (13) becomes, under the previous assumptions

$$\sigma(\mathbf{S})F(\mathbf{V} \mid \mathbf{S}) = \prod_{j=1}^{\infty} \frac{1}{2\pi\sigma_{a_j}\sqrt{N_0}} \exp\left[-\frac{1}{2}\frac{(a_j - b_j)^2}{N_0} - \frac{1}{2}\frac{(a_j - \mu_{a_j})^2}{\sigma_{a_j}^2}\right] \tag{17}$$

and

$$[\hat{\mathbf{S}}]_k = \frac{b_k\sigma_{a_k}^2 + \mu_{a_k}N_0}{\sigma_{a_k}^2 + N_0}. \tag{18}$$

Thus the b_k are unbiased estimators for this *a priori* distribution of $\mathbf{S}$. It is also to be noted that the b_k are easy to obtain via electronic instrumentation.[5] If $\sigma(\mathbf{S})$ and N_0 are known, the latter estimators are also easy to obtain.

E. M. Glaser
J. H. Park, Jr.
Radiation Lab.
The Johns Hopkins Univ.
Baltimore 2, Md.

[5] J. H. Park, Jr. and E. M. Glaser, "The extraction of waveform information by a delay line filter technique," 1957 IRE WESCON Convention Record, vol. 1, pt. 2, p. 171.

On Manasse, Price, and Lerner, "Loss of Signal Detectability in Band-Pass Limiters"*

Manasse, Price, and Lerner have recently demonstrated[1] the remarkable fact that it is possible to reduce the signal-detectability degradation factor of a band-pass limiter to unity by adding strong noise of low spectral intensity and sufficient bandwidth to its input. The purpose of this letter is to elucidate the phenomenon.

The input to the band-pass limiter consists of three components, which we may call the "input signal," the "original noise," and the "added noise." If the input consisted of the added noise alone, the output would be a broad, low spectrum of noise, putting negligible power into the band occupied by the original signal and noise. The effect of including the input signal and original noise is essentially to add these voltages, both attenuated equally but undistorted, to the output. Since the spectral intensity of the output due to the added noise can be made negligible compared to that due to the original noise within the band occupied by the latter by making the spectrum of the added noise sufficiently broad and low, it follows that the output signal-to-noise ratio in the band occupied by the original signal and the noise is the same as the original input signal-to-noise ratio; *i.e.*, there is no loss in detectability.

Of course, the same result can be obtained much more simply and with less attenuation by just omitting the limiter and not adding extra noise to the input nor filtering it out of the output. Thus, the effect of adding the extra noise, limiting, and filtering is just to attenuate the original input by something like the original-input-to-added-noise ratio. To see that this is the case, we refer to Blachman,[2] where the band-pass limiter is the special case $m = 1$, $n = 0$.

Although this paper does not so state its analysis applies to the case of a narrow-band signal as well as to the case of a sinusoidal signal, the output signal being defined as the ensemble-average output. Eq. (3) of Blachman,[2] which represents the total output power (signal plus noise), must be averaged over the distribution of the signal amplitude A if A is not constant. Eq. (4) represents the amplitude of the signal output; its square must be averaged over the distribution of A to obtain the output signal power if A is not constant. Thus, (5) gives the output signal-to-noise ratio if R is taken to be the ratio of the average input signal power to the average input noise power. [Incidentally, (5) of Blachman[3] has exactly the same form as (4) of Blachman[2] for the case $m = 1$, $n = 0$.]

For small input signal-to-noise ratios, (4) is proportional to A, and the phase of the output signal is always exactly equal to the phase of the input signal. Hence, for small signal-to-noise ratios, the output signal is an attenuated but undistorted version of the input signal. This conclusion can be applied to the case where strong noise is added to the input of a band-pass limiter by regarding the original input signal and noise as a new signal. Thus, this new signal (*i.e.*, the original input signal and noise) appears undistorted though attenuated in the output.

Nelson M. Blachman
Office of Naval Res.
London, England
On leave from Electronic
Defense Lab.
Mountain View, Calif.

* Received by the PGIT, July 15, 1958.

[1] R. Manasse, R. Price, and R. M. Lerner, "Loss of signal detectability in band-pass limiters," IRE Trans. on Information Theory, vol. IT-4, pp. 34–38; March, 1958.

[2] N. M. Blachman, "The output signal-to-noise ratio of a power-law device," *J. Appl. Phys.*, vol. 24, pp. 783–785; June, 1953.

[3] N. M. Blackman, "The demodulation of an F-M carrier and random noise by a limiter and discriminator," *J. Appl. Phys.*, vol. 20, pp. 38–47; January, 1949.

PGIT News

The Professional Group on Information Theory, in conjunction with the Professional Group on Circuit Theory, is sponsoring an International Symposium on Circuit and Information Theory, to be held at the University of California at Los Angeles on June 16–18, 1959. The purpose of the Symposium will be to consider recent advances in Information Theory and Circuit Theory, and in particular to explore areas of interest common to the two disciplines. A partial list of topics which are tentatively planned for the technical program are: application of linear graph theory to communication nets and circuits; switching circuits and coding; applications of matrix theory to circuit and information theory; specification and synthesis of matched filters; networks with random parameters; and characterization and optimization of nonlinear filters.

The organization of the Symposium will follow that of previous years. The Transactions of the Symposium will be published in advance, and adequate time will be allowed at the meeting for active participation and discussion from the floor. Every effort will be made to establish the atmosphere of a forum, rather than that of a lecture hall.

The deadline for submission of detailed 750-word summaries (in triplicate) of papers for presentation at the Symposium is December 22, 1958. In addition to papers which present new results, tutorial papers, especially those which expound advanced mathematical techniques of value to circuit and information theorists, will be considered.

All correspondence, including summaries, should be sent to: Dr. G. L. Turin, Hughes Research Laboratories, Culver City, Calif.

Contributors

Frederick J. Beutler (M'54) was born on October 3, 1926, in Berlin, Germany. He received the S.B. degree from the Massachusetts Institute of Technology, Cambridge, in 1949. In 1950 he returned to M.I.T. where he was granted the M.S. degree in 1951, and joined the staff of the M.I.T. Instrumentation Laboratory. Later he became a research engineer at North American Aviation.

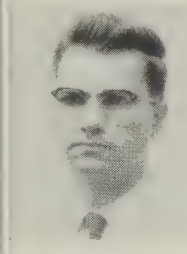
F. J. Beutler

In 1954, he entered the California Institute of Technology, Pasadena, where he received the Ph.D. degree in engineering science and mathematics in 1957. Since 1954 he has consulted widely in the fields of systems analysis, random processes, applied mathematics, and inertial navigation. He is presently on the faculty of the University of Michigan, Ann Arbor, where he is teaching in the area of instrumentation as an assistant professor of aeronautical engineering.

Dr. Beutler is a member of the American Mathematical Society, American Association for the Advancement of Science, and Sigma Xi.

❖

William M. Brown (M'57) was born on February 14, 1932, in Wheeling, W. Va. He received the B.S.E.E. degree in 1952 from the University of West Virginia, Morgantown, and the M.S.E. degree in 1955 and the D. Eng. degree in 1957 from The Johns Hopkins University, Baltimore, Md.



W. M. Brown

From 1952 to 1954, he worked in the Air Arm Division of Westinghouse. During 1954 to 1957 he was at The Johns Hopkins Radiation Laboratory, where he was a project leader in electronic countermeasures and also taught in the night school He then worked for a year at the Institute for Defense Analyses, Washington, D. C., and taught during the same period at George Washington University. He is now with the Electrical Engineering Department and Willow Run Laboratories, University of Michigan, Ann Arbor, Mich.

Dr. Brown is a member of the American Mathematical Society, Society for Industrial and Applied Mathematics, Eta Kappa Nu, Tau Beta Pi, Sigma Pi Sigma, and Sigma Xi.

❖

Sze-Hou Chang (S'46—A'48—SM'52) was born in Ningpo, China, on September 23, 1913. He received the B.S.E.E. degree from Chiao-Tung University in 1934. During 1934–1945, he taught at Tsing-Hua University, Hunan University, and Chiao-Tung University, all in China. He became a professor of electrical engineering at the latter university in 1943.

S.-H. Chang

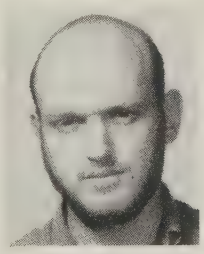
In 1945 he came to the United States and entered Harvard University, Cambridge, Mass., where he received the M.S. degree in 1946 and the Ph.D. degree in engineering sciences and applied physics in 1948.

He became associated with Northeastern University, Boston, Mass., in 1948 and is now a professor of research in communications. He has done research on speech analysis, speech coding for bandwidth compression, and communication theory.

Dr. Chang is a member of the Acoustical Society of America, Society of Industrial and Applied Mathematics, Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

❖

Michael B. Marcus was born in Brooklyn, N. Y., on March 5, 1936. He received the B.S. degree in electrical engineering from Princeton University, Princeton, N. J., in 1957, and, as a National Science Foundation Fellow, he received the M.S. degree in mathematics from Massachusetts Institute of Technology, Cambridge, Mass., in 1958.



M. B. Marcus

In 1956 he worked on computer logic systems for the Sandia Corporation, Albuquerque, N. M. As a consultant in mathematics for the Rand Corporation, Santa Monica, Calif., in 1957, he was concerned with decision processes and information theory.

Mr. Marcus was employed by the M. I. T. Instrumentation Laboratory before rejoining the Rand Corporation in 1958.

❖

J. A. McFadden, for a photograph and biography, please see page 60 of the March, 1958, issue of these Transactions.

❖

John N. Pierce (M'58) was born in Norwood, Mass., on July 9, 1933. He received the S.B. and S.M. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1955.

J. N. Pierce

Since 1952, Mr. Pierce has been employed at the U. S. Air Force Cambridge Research Center, Bedford, Mass., where he has worked on UHF data transmissions systems, long-range anti-jamming systems, and high-reliability scatter transmission.

At the present time, Mr. Pierce is engaged in studies of communication theory and reliability.

❖

Gerald E. Sacks was born in Brooklyn, N. Y., on March 22, 1933. During the period of 1953 to 1956 he served on active duty with the United States Army.

G. E. Sacks

He was conferred the B.S. and M.S. degrees in electrical engineering by Cornell University, Ithaca, N. Y., in 1958.

At present, he is working towards the Ph.D. degree in mathematics at Cornell University.

Mr. Sacks is a member of Eta Kappa Nu.

Index to

IRE TRANSACTIONS

ON

INFORMATION THEORY

Volume IT-4, 1958

IRE Transactions on Information Theory

Index to Volume IT-4, 1958

CONTENTS

Volume IT-4, Number 1, March, 1958

Volume IT-4, Number 2, June, 1958

Volume IT-4, Number 3, September, 1958

Volume IT-4, December, 1958, Number 4

Index to Authors

(Listed by month and page.)

Index to Subjects

(Listed by month and page.)

Index to Footnote References

The following is an alphabetical compilation of all footnote and bibliographical references published in PGIT Transactions *papers during the year. The month and page where each reference was listed are given in parentheses.*

A

Allen, F. E., "The general form of the orthogonal polynomials for simple series, with proofs of their simple properties," *Proc. Roy. Soc. Edinburgh*, vol. 50, pp. 310–320; 1935. (Mar 25)
Amiantov, I. N. and Tikhonov, V. I., "The effect of normal fluctuations on typical nonlinear elements," *Bull. Acad. Sci. USSR*, pp. 33–42; April, 1956. (Jun 71)
Anderson, R. L. and Hauseman, E. E., "Tables of Orthogonal Polynomial Values Extended to N = 104," Agricultural Exper. Station, Iowa State College of Agricultural and Mechanical Arts, Ames, Ia., Res. Bull. 297; April, 1942. (Mar 25)
Anderson, T. W., "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proc. Amer. Math. Soc.*, vol. 6, pp. 170–176; 1955. (Sep 126)
Arthur, G. R., "Statistical properties of output of a frequency sensitive device," *J. Appl. Phys.*, vol. 25, pp. 1185–1195; September, 1954. (Jun 83)

B

Bailey, A. E., "Integration in pulse radar systems," in "Communication Theory," papers of 1952 London Symposium, W. Jackson, ed., Academic Press, Inc., New York, N. Y.; 1953. (Sep 114)
Balakrishan, A. V. and Drenick, R., "On optimum nonlinear extraction and coding filters," IRE Trans. on Information Theory, vol. IT-2, pp. 166–172; September, 1956. (Jun 88)
Barrett, J. F. and Lampard, D. G., "An expansion for some second-order probability distributions and its application to noise problems," IRE Trans. on Information Theory, vol. IT-1, pp. 10–15; March, 1955. (Jun 71) (Jun 73) (Dec 172)
Bartlett, M. S., "An Introduction to Stochastic Processes," Cambridge University Press, Cambridge, Eng.; 1955. (Mar 17) (Dec 162)
Bateman, H., "Higher Transcendental Functions," McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1; 1953. (Mar 51)
Battin, R. H.: See Laning, J. H., Jr.
Baum, R. F., "The correlation function of smoothly limited Gaussian noise," IRE Trans. on Information Theory, vol. IT-3, pp. 193–197; September, 1957. (Jun 72)
Baxter, G., "A strong limit theorem for Gaussian processes," *Proc. Amer. Math. Soc.*, vol. 7, pp. 522–527; June, 1956. (Jun 66)
Bedrosian, E., "Weighted PCM," IRE Trans. on Information Theory, vol. IT-4, p. 45 (Mar 58)
Bellman, R., "Dynamic Programming," Princeton University Press, Princeton, N. J.; 1957. (Mar 58) (Dec 151)
———, "Functional Equations in the Theory of Dynamic Programming-VII: An Integro-Differential Equation for the Fredholm Resolvent," The RAND Corp., Santa Monica, Calif., Paper P-859; May 7, 1956. (Mar 5)
———, and Kalaba, R., "Dynamic programming and statistical communication theory," *Proc. Natl. Acad. Sci.*, vol. 43, pp. 749–751; August, 1957. (Dec 151)
———, and ———, "On the role of dynamic programming in statistical communication theory," IRE Trans. on Information Theory, vol. IT-3, pp. 197–203; September, 1957. (Dec 151)
Bennett, W. R., "Spectra of quantized signals," *Bell Sys. Tech. J.*, vol. 27, pp. 446–472; July, 1948. (Mar 49)
Bernstein, R. I.: See Miller, K. S.
Beutler, F. J., "A generalization of Wiener optimum filtering and prediction," Doctoral dissertation, California Inst. Tech., Pasadena, Calif.; 1957. (Dec 167)
Birdsall, T. G.: See Peterson, W. W.
Blachman, N. M., "The demodulation of an F-M carrier and random noise by a limiter and discriminator," *J. Appl. Phys.*, vol. 20, pp. 38–47; January, 1949. (Dec 174)
———, "The output signal-to-noise ratio of a power-law device," *J. Appl. Phys.*, vol. 24, pp. 783–785; June, 1953. (Dec 174)
Black, H. S., "Modulation Theory," D. Van Nostrand Co., Inc., New York, N. Y.; 1953. (Mar 49)
Blanc-Lapierre, A. and Fortet, R., "Theorie Des Functions Aleatories," Masson and Co., Paris, France; 1953. (Dec 144)
Blasbalg, H., "Bernoulli Sequential Filtering," Radiation Lab., The Johns Hopkins University, Baltimore, Md., Tech. Rep. No. AF-27; January, 1956. (Dec 151)
Bloom, F., Chang, S., Harris, B., Hauptschein, A., and Morgan, K., "Improvement of binary transmission by null-zone reception," Proc. IRE, vol. 45, pp. 963–975; July, 1957. (Mar 59) (Dec 151)
Blum, M., "An extension of the minimum mean square prediction theory for sampled input data," IRE Trans. on Information Theory. vol. IT-2, pp. 176–184; September, 1956. (Mar 28)
———, "Fixed memory least squares filters using recursion methods," IRE Trans. on Information Theory, vol. IT-3, pp. 178–182; September, 1957. (Mar 24)
Blumenson, L. E.: See Miller, K. S.
Bond, F. E., "The information-bearing elements in bandwidth limited signals," Doctoral dissertation submitted to Polytechnic Inst. of Brooklyn, Brooklyn, N. Y.; June, 1956. (Sep 112)
Brown, J. L., Jr., "On a cross-correlation property for stationary random processes," IRE Trans. on Information Theory, vol. IT-3, pp. 28–31; March, 1957. (Jun 71) (Jun 73) (Dec 172)
Brown, W. M., "Noise Statistics After Transformations Commonly Found in Circuits," Dept. of Commerce, Office of Tech, Services, Washington, D. C., OTS No. PB121515; 1955. (Dec 144)
Bunimovich, V. I., "Amplitude peaks of random noise," *Zh. Tekh. Fiz.*, vol. 21, pp. 625–636; June, 1951. (In Russian.) (Mar 31)
———, "The fluctuating process as an oscillation with random amplitude and phase," *Zh. Tekh. Fiz.*, vol. 19, pp. 1231–1259; November, 1949. (In Russian.) (Mar 33)
———, "Fluctuation processes as oscillations of random amplitude and phase," *Zh. Tekh. Fiz.* (*Leningrad*), vol. 19, pp. 1231–1259; November, 1949. (Mar 35)
Bussgang, J. J., "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rep. 216; March 26, 1952. (Jun 71) (Jun 73)

——— ———, and Middleton, D., "Optimum sequential detection of signals in noise," IRE Trans. on Information Theory, vol. IT-1, pp. 5–18; December, 1955. (Dec 151)

C

Campbell, G. A., "Probability curves showing Poisson's exponential summation," *Bell Sys. Tech. J.*, vol. 2, pp. 95–113; 1923. (Mar 39)
Chang, S. S. L., See Bloom, F. J.
———, *et al.*, "Supplementary Notes on Evaluation and Optimization of Digital Communication Systems," College of Eng., New York University, New York, N. Y., First Sci. Rep., AFCRC-TN-57-355, AD117047; January 15, 1957 to April 15, 1957. (Mar 49)
Cheatham, T. P.: See Lee, Y. W.
Cohen, R., "Some Analytical and Practical Aspects of Wiener's Theory of Predictions," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rep. No. 69; June 2, 1948. (Jun 71)
Cohn, G. I. and Flesher, G. T., "Investigation Pertaining to Elimination of Ambiguities Due to High Pulse Repetition Rates," Sixth Quart. Prog. Rep., Signal Corps Contract No. DA-36-039 SC-56696; July 19, 1955. (Sep 114)
Cooper, G. R., "System measurement with binary noise," to be published. (Dec 160)
Courant, R., "Differential and Integral Calculus," Nordeman Publishing Co., Inc., New York, N. Y., vol. 2; 1937. (Mar 49)
———, and Hilbert, D., "Methoden der Mathematischen Physik," Springer Verlag, Berlin, Ger., vol. 1; 1931. (Mar 7)
Cramer, H., "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J.; 1946. (Mar 31) (Jun 69) (Jun 85) (Sep 26)
———, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J., 1951. (Mar 50) (Dec 144)
Croney, J., "Clutter on radar displays," *Wireless Eng.*, vol. 33, pp. 83–96; April, 1956. (Mar 50)

D

Darling, D. A. and Siegert, A. J. F., "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena—Part I," The RAND Corp., Santa Monica, Calif., Paper P-738; September 10, 1955. Also IRE Trans. on Information Theory, vol. IT-3, pp. 32–37; March, 1957. (Mar 4)
Davenport, W. B., Jr., "Signal-to-noise ratios in band-pass limiters," *J. Appl. Phys.*, vol. 24, pp. 720–727; June, 1953. (Mar 35)
———, and Root, W. L., "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Book Co., Inc., New York, N. Y.; 1958. (Sep 125)
Davis, H., "Radar problems and information theory," 1953 IRE Convention Record, pt. 8, pp. 39–45. (Sep 114)
Doob, J. L., "The Brownian movement and stochastic equations," *Ann. Math.*, vol. 43, pp. 351–369; April, 1942. (Dec 166)
———, "Stochastic Processes," John Wiley and Sons, Inc., New York, N. Y.; 1953. (Mar 21) (Jun 66) (Jun 85) (Jun 88) (Dec 144) (Dec 167)
Drenick, R.: See Balakrishan, A. V.
Dwork, B. M., "Detection of a pulse superimposed on fluctuation noise," Proc. IRE, vol. 38, pp. 771–774; July, 1950. (Sep 114)

E

Elias, P., "Coding for Two Noisy Channels," in "Information Theory," C. Cherry, ed., Thornton Butterworth, Ltd., London, Eng.; 1956. (Dec 152)
Erdelyi, A. "Higher Transcendental Functions," McGraw-Hill Book Co., Inc., New York, N. Y., vol. 2; 1953. (Jun 75)

F

Fano, R. M., "Signal-to-Noise Ratio in Correlation Detectors," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rep. No. 186; February 19, 1951. (Sep 114)
Favreau, R. R., Low, H., and Pfeffer, I., "Evaluation of complex statistical functions by an analog computer," 1956 IRE Convention Record, pt. 4, pp. 31–37. (Mar 19)
Feller, W., "An Introduction to Probability Theory and its Applications," John Wiley and Sons, Inc., New York, N. Y., vol. 1; 1950. (Mar 17) (Dec 160)
Flesher, G. T.; See Cohn, G. I.
Fogel, L.: See Jagerman, D. L.
Fortet, R.: See Blanc-Lapierre, A.
Fox, L. and Goodwin, E. T., "The numerical solution of non-singular linear integral equations," *Phil. Trans. Roy. Soc. London*, vol. A245, pp. 501–534; February, 1953. (Mar 19)
Fox, W. C.: See Peterson, W. W.
Freedman, J. and Margolin, J., "Signal-to-Noise Improvement through Integration in a Delay-Line Filter System," Lincoln Lab., M.I.T., Lexington, Mass., Tech. Rep. No. 22; May 13; 1953. (Sep 114)
Freund, J. E.: See Miller, I.

G

Gabor, D., *J. IEE*, pt. 3, vol. 93, p. 429; 1946. (Mar 53)
George, S. F., "Time Domain Correlation Detectors vs Conventional Frequency Domain Detectors," Naval Res. Lab., Washington, D. C., Rep. 4332; May, 1954. (Sep 121)
———, and Zamanakos, A. S., "Comb filters for pulsed radar use," Proc. IRE, vol. 42, pp. 1159–1165; July, 1954. (Sep 114)
Gilbert, E., "A comparison of signaling alphabets," *Bell Sys. Tech. J.*, vol. 31, pp. 504–522; May 1952. (Dec 146)
Glaser, E. M.: See Park, J. H., Jr.
Glovazky, A., "Determination of redundancies in a set of patterns," IRE Trans. on Information Theory, vol. IT-2, pp. 151–153; December, 1956. (Sep 127)
Golay, M. J. E., "Binary coding," IRE Trans. on Information Theory, vol. IT-4, pp. 23–38; September, 1954. (Sep 103) (Dec 145)
———, "Notes on digital coding," Proc. IRE, vol. 37, p. 657; June, 1949. (Jun 78) (Sep 103)
Goodwin, E. T.: See Fox, L.
Goursat, E., "Cours d'Analyse," Paris, France, 2nd ed., vol. 3; 1915. (Sep 129)
Graves, R. E., "Design and Analysis of an Optimum Comb Filter Based upon an Information Storage Device," Goodyear Aircraft Corp., Akron, Ohio, Rep. GERA-20; January, 1953. (Sep 114)
———, "Miscellaneous Considerations in Comb Filter Design," Goodyear Aircraft Corp., Akron, Ohio, Rep. GERA-24; August, 1954. (Sep 114)
Grenander, U., "Stochastic processes and statistical inference," *Arkiv. Matematik*, band 1, no. 17, theorem, p. 257; 1950. (Jun 66)
Grobner, W. and Hofreiter, N., "Integraltafel," Springer-Verlag Co., Inc., Vienna, Austria; 1950. (Dec 144)
Guillemin, E. A., "Communication Networks," John Wiley and Sons, Inc., New York, N. Y., vol. 2; 1949. (Sep 123)

H

Hamming, R. W., "Error detecting and error-correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; January, 1950. (Jun 77) (Sep 103) (Dec 145) (Dec 152)
Harrington, J. V., "Storage of small signals on a dielectric surface," *J. Appl. Phys.*, vol. 21, pp. 1048–1053; October, 1950. (Sep 114)
———, and Rogers, T. F., "Signal-to-noise improvement through integration in a storage tube," Proc. IRE, vol. 38, pp. 1197–1203; October, 1950. (Sep 114)
Harris, B.: See Bloom, F. J.
Hauptschein, A.: See Bloom, F. J.
Hauseman, E. E.: See Anderson, R. L.
Hilbert, D.: See Courant, R.
Hlrich, W., "Non-binary error correction codes," *Bell Sys. Tech. J.*, vol. 36, pp. 1341–1388; November, 1957. (Jun 77)
Hoffman, W. C., "The joint distribution of n successive outputs of a linear detector," *J. Appl. Phys.*, vol. 25, pp. 1006–1007; August, 1954. (Dec 163)
Hofreiter, N.: See Grobner, W.
Hopper, A. L.: See Pierce, J. R.
Huggins, W. H., "Signal-flow graphs and random signals," Proc. IRE, vol. 45, pp. 74–86; January, 1957. (Mar 17)
———, "Signal theory," IRE Trans. on Circuit Theory, vol. CT-3, pp. 210–216; December, 1956, (Dec 173)

J

Jackson, D., "Fourier Series and Orthogonal Polynomials," Math. Assoc. of America; 1941. (Dec 172)
Jagerman, D. L. and Fogel, L., "Some general aspects of the sampling theorem," IRE Trans. on Information Theory, vol. IT-2, pp. 139–146; December, 1956. (Sep 110)

Jahnke, E. and Emde, F., "Tables of Functions," Dover Publications, Inc., New York, N. Y.; 1945. (Mar 51)
James, H. M., Nichols, N. B., and Phillips, R. S., "Theory of Servomechanisms," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 25; 1947. (Mar 26) (Mar 54) (Dec 160) (Dec 170)
Jordan, C., "Calculus of Finite Differences," Chelsea Publishing Co., New York, N. Y.; 1950. (Mar 26)

K

Kac, M. and Siegert, A. J. F., "On the theory of noise in radio receivers with square law detectors," *J. Appl. Phys.*, vol. 18, pp. 396–397; April, 1947. (Dec 163)
Kahn, H., "Application of Monte Carlo," RAND Corp., Santa Monica, Calif., Res. Memo. RM-1237-AEC; April 19, 1954. Revised, April 27, 1956. (Mar 28)
Kalaba, R.: See Bellman, R.
Kelly, J., "A new interpretation of information rate," *Bell Sys. Tech. J.*, vol. 35, pp. 917–926; July, 1956. (Dec 151)
Khinchin, A. I., "Statistical Mechanics," Dover Publications, Inc., New York, N. Y.; 1949. (Dec 144)
Kohlenberg, A., "Notes on the Zero Distribution of Gaussian Noise," M.I.T., Lincoln Lab., Lexington, Mass., Tech. Memo. 44; October, 1953. (Mar 15)
Kolmogoroff, A. N., "Foundations of the Theory of Probability," Chelsea Publishing Co., Inc., New York, N. Y.; 1950. (Dec 144)
————, "Interpolation and extrapolation von stationären zufalligen Folgen," *Bull. Acad. Sci. U. S. S. R., Ser. Math.*, vol. 5, pp. 5–14; 1941. (Dec 169)

L

Lampard, D. G.: See Barrett, J. F.
————, "The probability distribution for the filtered output of a multiplier whose inputs are correlated, stationary, Gaussian time-series," IRE Trans. on Information Theory, vol. IT-2, pp. 4–11; March, 1956. (Mar 4)
Landan, E. and Ostrowski, A., "On the diophantine equation $ay^2 + by + c = dx^n$," *Proc. London Math. Soc.*, ser. 2, vol. 19, pp. 276–280; 1921. (Jun 79)
Laning, J. H., Jr., and Battin, R. H., "Random Processes in Automatic Control," McGraw-Hill Book Co., Inc., New York, N. Y.; 1956. (Jun 72) (Jun 86) (Sep 125) (Dec 144) (Dec 159)
Lawson, J. L. and Uhlenbeck, G. E., "Threshold Signals," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y.; 1950. (Mar 17) (Jun 71) (Sep 114) (Dec 144)
Lee, Y. W., Cheatham, T. P., and Wiesner, J. B., "Application of correlation analysis to the detection of periodic signals in noise," Proc. IRE, vol. 38, pp. 1165–1171; October, 1950. (Sep 114)
Lerner, R. M.: See Manasse, R.
Levenson, N., "Gap and Density Theorems," American Mathematical Society Colloquium Publication, New York, N. Y., vol. 26; 1940. (Sep 110)
————, "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," *J. Math. Phys.*, vol. 26, pp. 110–119; July, 1947. (Dec 169)
Lewis, T., "99.9 and 0.1% points of the x^2 distribution," *Biometrika*, vol. 40, pts. 3–4, pp. 421–426; December, 1953. (Mar 39)
Licklider, J. C. R. and Pollack, I., "Effects of differentiation, integration, and peak clipping upon the intelligibility of speech," *J. Acoust. Soc. Amer.*, vol. 20, pp. 42–51; January, 1958. (Sep 111)
Lloyd, S. P., "Binary block coding," *Bell Sys. Tech. J.*, vol. 36, pp. 517–535; March, 1957. (Jun 78)
Loeve, M., "Probability Thoery," D. Van Nostrand Co., Inc., New York, N. Y.; 1955. (Jun 66) (Dec 144)
Longuet-Higgins, M. S., "On the intervals between successive zeros of a random function," *Proc. Roy. Soc. (London) A*, vol. 246, pp. 99–118; July, 1958. (Dec 162)
Low, H.: See Favreau, R. R.

M

Magness, T. A., "Spectral responses of a quadratic device," *J. Appl. Phys.*, vol. 25, pp. 1357–1365; November, 1954. (Dec 160)
Manasse, R., Price, R., and Lerner, R. M., "Loss of signal detectability in band-pass limiters," IRE Trans. on Information Theory, vol. IT-4, pp. 34–38; March, 1958. (Dec 174)
Marcum, J. I., "A Statistical Theory of Target Detection by Pulsed Radar," USAF, Project RAND, RAND Corp., Santa Monica, Calif., RM-754; 1947. (Mar 50)
————, "A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix," USAF, Project RAND, RAND Corp., Santa Monica, Calif., RM-753, p. 11; April 25, 1952. (Mar 38)
Margolin, J.: See Freedman, J.
Marton, L., "Advances in Electronics," Academic Press, Inc., New York, N. Y., vol. 3; 1951. (Mar 49)
McFadden, J. A., "An approximation for the symmetric quadrivariate normal integral," *Biometrika*, vol. 43, pp. 206–207; June 1956. (Dec 159)
————, "The axis-crossing intervals of random functions," IRE Trans. on Information Theory, vol. IT-2, pp. 146–150; December, 1956. (Mar 14)
————, "The axis-crossing intervals of random functions II," IRE Trans. on Information Theory, vol. IT-4, pp. 14–24; March, 1958. (Dec 159)
————, "Urn models of correlation and a comparison with the multivariate normal integral," *Ann. Math. Stat.*, vol. 26, pp. 478–489; September, 1955. (Mar 23) (Dec 159)
————, "The variance of zero-crossing intervals," *Ann Math. Stat.*, vol. 28, p. 529; June, 1957. (Mar 21)
Mehler, "Higher Transcendental Functions," McGraw-Hill Book Co., Inc., New York, N. Y., vol. 2; 1953. (Jun 71)
Middleton, D.: See Bussgang, J. J.
————, "On the detection of stochastic signals in additive normal noise—Part I," IRE Trans. on Information Theory, vol. IT-3, pp. 86–121; June, 1957. (Jun 65)
————, "Spurious signals caused by noise in triggered circuits," *J. Appl. Phys.*, vol. 19, pp. 817–830; September, 1948. (Mar 31)
————, and Van Meter, D., "Detection and extraction of signals in noise from point of view of statistical decision theory, parts I and II," *J. Soc. Ind. Appl. Math.*, vol. 3, pp. 192–253; December, 1955, and vol. 4, pp. 86–119; June, 1956. (Dec 144) (Dec 173)
Miller, I. and Freund, J. E., "Some Distribution Theory Connected with Gaussian Processes," Dept. of Statist., Virginia Polytech. Inst., Blacksburg, Va., Tech. Rep. 19; May 1956. (Mar 17)
Miller, K. S. and Bernstein, R. I., "Generalized Rayleigh processes," (abstract), *Bull. Amer. Math. Soc.*, vol. 63, pp. 196–197; May, 1957. (Mar 31)
————, ————, and Blumenson, L. E., "Generalized Rayleigh processes," *Quart. Appl. Math.*, to be published. (Mar 31)
Morgan, K. C.: See Bloom, F. J.
Munroe, M. E., "Introduction to Measure and Integration," Addison-Wesely Co., Cambridge, Mass. 1953. (Dec 144)
Muroga, S., "On the capacity of a discrete channel I," *J. Phys. Soc., Japan*, vol. 8, pp. 484–494; July-August, 1953. (Dec 152)

N

Nagell, T., "Introduction to Number Theory," John Wiley and Sons, Inc., New York, N. Y.; 1951. (Sep 108)
————, "Sur l' impossibilité de quelques équations à deux imdéterminées," *Norsk. Matemat. Foren. Skrifter*, ser. 1, no. 13; 1923. (Jun 79)
Nichols, N. B.: See James, H. M.
North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Labs., Rep. PTR-6C; June, 1943. (Sep 114)
Nuttall, A. H., "Invariance of Correlation Functions under Nonlinear Transformations," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Quart. Progress Rep.; October 15, 1957. (Jun 71)

O

Oliver, B. M., Pierce, J. R., and Shannon, C. E., "The Philosophy of PCM," Proc. IRE, vol. 36, pp. 1324–1331; November, 1948. (Mar 49)
Orr, L. W., "Wideband Amplitude Distribution Analyzer of Voltage Sources," Electronics Defense Group, Univ. of Michigan, Ann Arbor, Mich., Tech. Rep. No. 22; 1953. (Dec 144)
Ostrowski, A.: See Landan, E.

P

Paley, R. E. A. C. and Wiener, N., "Fourier Transforms in the Complex Domain," American Mathematical Society Colloquium Publication, New York, N. Y., vol. 19, 1934. (Sep 110) (Dec 169)

Palmer, D. S., "Properties of random functions," *Proc. Camb. Phil. Soc.*, vol. 52, pp. 672–686; October, 1956. (Mar 17) (Dec 161)
Park, J. H., Jr. and Glaser, E. M., "The extraction of waveform information by a delay line filter technique," 1957 WESCON Convention Record, pt. 2, pp. 171–184. (Dec 174)
Pearson, K., "Tables of the Incomplete Gamma Function," Cambridge Press, New York, N. Y.; 1951. (Mar 39)
Peterson, W. W., Birdsall, T. G., and Fox, W. C., "The theory of signal detectability," IRE Trans. on Information Theory, vol. IT-4, pp. 171–212; September, 1954. (Mar 34)
Pfeffer, I.: See Favreau, R. R.
Phillips, R. S.: See James, H. M.
Pierce, J. R.: See Oliver, B. M.
————, and Hopper, A. L., "Non-synchronous time division with holding and random sampling," Proc. IRE, vol. 40, pp. 1079–1088; September, 1942. (Sep 110)
Plackett, R. L., "A reduction formula for multivariate normal integrals," *Biometrika*, vol. 41, pp. 351–360; December, 1954. (Dec 162)
Plotkin, M., "Binary Codes with Specified Minimum Distance," Moore School of Elec. Eng., Univ. of Pennsylvania, Philadelphia, Pa.; 1952. (Jun 77)
Pollack, I.: See Licklider, J. C. R.
Price, R.: See Manasse, R.
————, "A note on the envelope and phase-modulated components of narrow-band Gaussian noise," IRE Trans. on Information Theory, vol. IT-1, pp. 9–13; September, 1955. (Mar 35)

R

Ragazzini, J. R.: See Zadeh, L. A.
Ramakrishna, B. S., *et al.*, "Statistical studies in some Indian languages with application to communication engineering," *J. Inst. Telecomm. Engrs. (India)*, vol. 4, pp. 25–35; December, 1957. (Sep 127)
Reed, I. S., "A class-of multiple-error-correcting codes and the decoding scheme," IRE Trans. on Information Theory, vol. IT-4, pp. 38–49; September, 1954. (Dec 146) (Dec 152)
Reeves, A. H., U. S. Patent 2,272,070, assigned to Internatl. Standard Electric Corp.; February 3, 1942. Also, French Patent No. 852,183; October 23, 1939. (Mar 49)
Rice, S. O., "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 306–310; July, 1944. (Mar 15) (Mar 53) (Jun 69) (Jun 85) (Sep 100) (Dec 160)
————, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46–156; January, 1945. (Mar 15) (Mar 37) (Mar 50) (Jun 69) (Jun 86) (Sep 100) (Sep 111) (Dec 160) (Dec 163)
————, "Mathematical analysis of random noise," reprinted in "Selected Papers on Noise and Stochastic Processes," N. Wax, ed., Dover Publications, Inc., New York, N. Y.; 1954. (Mar 30)
————, "Properties of a sine wave plus random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 109–157; January, 1948. (Mar 31)
————, "Statistical properties of a sine wave plus random noise," *Bell Sys. Tech. J.*, vol. 27, p. 129; January, 1948. (Jun 83)
Robin, L., "The autocorrelation function and power spectrum of clipped thermal noise. Filtering of simple periodic signals in this noise," *Ann. Telecomm.*, vol. 7, pp. 375–387; September, 1952. (Jun 72)
Rogers, T. F.: See Harrington, J. V.
Root, W. L.: See Davenport, W. B., Jr.

S

Schultheiss, P. M.: See Steinberg, H.
————, Wogrin, A., and Zweig, F., "Short-time frequency measurement of narrow-band random signals in the presence of wideband noise," *J. Appl. Phys.*, vol. 25, pp. 1025–1036; August, 1954. (Jun 83)
Schwartz, L., "Theorie des Distributions," Hermann, Paris, France, 1950. (Jun 84)
Shannon, C. E.: See Oliver, B. M.
————, "Communication in the presence of noise," Proc. IRE, vol. 37, pp. 10–21; January, 1949. (Sep 129)
————, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423; July, 1948. (Mar 49) (Sep 103) (Sep 127) (Dec 151)
————, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 623–656; October, 1948. (Mar 49) (Sep 110) (Dec 151)
————, and Weaver, W., "The Mathematical Theory of Communications," University of Illinois Press, Urbana, Ill.; 1949. (Dec 152)
Shapiro, H. S. and Slotnick, D. L., "On the Mathematical theory of error-correcting codes," *IBM J. Res.*, to be published. (Jun 78)
Sherman, S., "A theorem of convex set with applications," *Ann. Amer. Math. Stat.*, vol. 26, pp. 763–767; 1955. (Sep 126)
Siegert, A. J. F.: See Darling, D. A., and Kac, M.
————, "Passage of stationary processes through linear and non-linear devices," IRE Trans. on Information Theory, vol. IT-3, pp. 4–25; March, 1954. (Mar 4)
————, "On the Roots of Markoffian Random Functions," The RAND Corp., Santa Monica, Calif., Rep. RM-447; September, 1950. (Mar 17)
————, "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena—Part II, Examples," The RAND Corp., Paper P-730; September, 1955. Also, IRE Trans. on Information Theory, vol. IT-3, pp. 38–43; March, 1957. (Mar 4)
Silverman, R. A., "On binary channels and their cascades," IRE Trans. on Information Theory, vol. IT-1, pp. 19–27; December, 1955. (Dec 152)
Slepian, D., "A class of binary signaling alphabets," *Bell Sys. Tech. J.*, vol. 35, pp. 203–234; January, 1956. (Dec 145) (Dec 152)
Slotnick, D. L.: See Shapiro, H. S.
Steinberg, H., Schultheiss, P. M., Wogrin, C. A., and Zweig, F., "Short-time frequency measurement of narrow-band random signals by means of a zero-counting process," *J. Appl. Phys.*, vol. 26, pp. 195–201; February, 1955. (Mar 16) (Jun 83)

T

Teichroew, D., "A table of millile probability points of the incomplete gamma distribution," Natl. Bur. Standards, Washington, D.C.; February, 1954. (Mar 39)
ter Haar, D., "Elements of Statistical Mechanics," Rinehart and Co., Inc., New York, N. Y.; 1954. (Dec 144)
Tikhonov, V. I.: See Amiantov, I. N.
Titchmarsh, E. C., "Introduction to the Theory of Fourier Integrals," Oxford University Press, New York, N. Y.; 1937. (Mar 53)
————, "The zeros of certain integral functions," *Proc. Lond. Math. Soc.*, vol. 25, pp. 283–302; May 14, 1926. (Sep 111)
Truxal, J. G., "Automatic Feedback Control System Synthesis," McGraw-Hill Book Co., Inc., New York, N. Y.; 1955. (Sep 125)

U

Uhlenbeck, G. E.: See Lawson, J. L., and Wang, M. C.

V

Van Meter, D.: See Middleton, D.
Ville, J. A., *Câbles & Transm.*, no. 1, p. 16; 1948. (Mar 53)

W

Wald, A., "Sequential Analysis," John Wiley and Sons, Inc., New York, N. Y.; 1947. (Dec 151)
Wang, M. C. and Uhlenbeck, G. E., "On the theory of the Brownian Motion II," *Rev. Mod. Phys.*, vol. 17, pp. 323–342; April–July, 1945. (Mar 5) (Dec 163)
Watson, G. N.: See Whittaker, E. T.
————, "A Treatise on the Theory of Bessel Functions," Cambridge University Press, Cambridge, Eng.; 1944. (Mar 13)
Wax, N., ed., "Noise and Stochastic Processes," Dover Publications, Inc., New York, N. Y.; 1949. (Dec 144)
Weaver, W.: See Shannon, C. E.
Whittaker, E. T. and Watson, G. N., "Modern Analysis," The MacMillan Co., New York, N. Y.; 1947. (Jun 67)
Widder, D. V., "The Laplace Transform," Princeton University Press, Princeton, N. J.; 1946. (Jun 69)
Wiener, N.: See Paley, R. E. A. C.
————, "Extrapolation, Interpolation and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, N. Y.; 1949. (Sep 127) (Dec 144)
————, "Generalized harmonic analysis," *Acta. Math.*, vol. 55, pp. 117–258; 1930. (Jun 85)
Wiesner, J. B.: See Lee, Y. W.

Winter, D. F., "A Gaussian noise generator for frequencies down to zero cycles per second," 1954 IRE Convention Record, pt. 4, pp. 23–29. (Dec 144)
Wishart, J., "x^2 probabilities for large numbers of degrees of freedom," *Biometrica*, vol. 43, pts. 1–2, pp. 92–95; June, 1956. (Mar 38)
Wogrin, C. A.: See Schultheiss, P. M., and Steinberg, H.
Woodward, P. M., "Probability and Information Theory," McGraw-Hill Book Co., Inc., New York, N. Y.; 1953. (Mar 53)

Y

Yen, J. L., "On the non-uniform sampling of bandwidth-limited signals," IRE Trans. on Circuit Theory, vol. CT-3, pp. 251–257; December, 1956. (Sep 110)

Z

Zadeh, L. A. and Ragazzini, J. R., "Optimum filters for the detection of signals in noise," Proc. IRE, vol. 40, pp. 1123–1231; October, 1952. (Sep 114)
Zamanakos, A. S.: See George, S. F.
Zaremba, S. K., "Covering problems concerning abelian groups," *J. Lond. Math. Soc.*, vol. 27, pp. 242–246; April, 1952. (Sep 103)
Zweig, F.: See Schultheiss, P. M., and Steinberg, H.

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT TRANSACTIONS. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the PROCEEDINGS OF THE IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE TRANSACTIONS ON INFORMATION THEORY is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. In addition, a period of approximately two months is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates. In addition, the IRE NATIONAL CONVENTION RECORD is published in July, and the IRE WESCON CONVENTION RECORD in the Fall. A bound collection of Information Theory papers delivered at these conventions is mailed gratis to all PGIT members.

All technical manuscripts and editorial correspondence should be addressed to George A. Deschamps, University of Illinois, Urbana, Ill. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Laurin G. Fischer, ITT Laboratories, 492 River Road, Nutley 10, N. J.

INSTITUTIONAL LISTINGS

The IRE Professional Group on Information Theory is grateful for the assistance given by the firms listed below and invites application for Institutional Listing from other firms interested in the field of Information Theory.

MOTOROLA, INC., 4545 West Augusta Blvd., Chicago 51, Ill.
Television, Home & Auto Radio, Phonograph & Hi-Fi, Communications & Industrial Electronics

THE RAMO-WOOLDRIDGE CORPORATION
5500 West El Segundo Blvd., Los Angeles 45, Calif.

REPUBLIC AVIATION CORP., Farmingdale, N. Y.
Aircraft, Missiles, Drones, Electronic Analyzers; U. S. Distr. of Alouette Turbine-Powered Helicopter